Earnshaw, Katherine

A forensic phonetic investigation of regional variation and accommodation in West Yorkshire

University of Huddersfield

A forensic phonetic investigation of regional variation and accommodation in West Yorkshire

Katherine Elizabeth Earnshaw

A thesis submitted to the University of Huddersfield in fulfilment of the requirements of the degree of Doctor of Philosophy

December 2020

Abstract

This thesis presents an examination of regional variation and speech accommodation in two socially salient features of West Yorkshire English. The first aim of this research is to consider the extent to which local level variation exists across the West Yorkshire boroughs of Bradford, Kirklees and Wakefield. The second aim is to evaluate the effects of speech accommodation, the process whereby speakers adapt their speech production according to whom to they are talking (Giles, 1973; Giles & Powesland, 1975; Trudgill, 1981), in forensically-relevant contexts. The findings from these examinations inform how generalisable population data is for West Yorkshire across the three boroughs and also demonstrate to what extent accommodation could impact forensic speaker comparison (FSC) casework.

The specific features examined in this thesis are the West Yorkshire FACE vowel and word-medial, intervocalic /t/. The motivations for examining these variables are twofold. Firstly, previous investigations of West Yorkshire English have suggested that these variables may be realised in different ways across the region and secondly, both variables appear to be socially salient in the speech community under investigation. As speech accommodation has been found to occur more often and to a stronger degree with respect to features that are socially salient (Cao, 2018; Smith & Holmes-Elliott, 2015; Trudgill, 1986), it was expected that the participants in this investigation would accommodate in respect of these speech parameters. However, the main focus of this investigation is to examine the magnitude and direction of any accommodation behaviour, and to evaluate the potential consequences this may have for FSC outcomes.

This study is one of the first to make use of the newly published West Yorkshire Regional English Database (WYRED; Gold, Ross, & Earnshaw, 2018). The study analyses the speech of 30 males from West Yorkshire recorded completing three semi-spontaneous speaking tasks that utilise different interlocutors. Participants are equally split across the boroughs of Bradford, Kirklees and Wakefield and form a homogenous population in terms of age, gender and language background, enabling a systematic evaluation of regional variation. For the analysis of FACE, measurements are taken of the first three formants at 25%, 50% and 75%

across the total vowel duration. Using these measurements, a series of statistical analyses are conducted in order to establish levels of variability across boroughs and across tasks. Additionally, realisations of intervocalic /t/ are analysed auditorily and assessments of variability between boroughs are carried out as well as an examination of changes in T-glottaling rates across tasks. For both speech parameters, accommodation is evaluated using an acoustic-phonetic approach whereby the participants' realisations are considered in relation to those of their respective interlocutors.

The findings of the investigations presented in this thesis reveal that FACE productions vary at the local borough level, specifically in terms of midpoint F2 values, whereas /t/ productions are not regionally stratified across West Yorkshire. Based on these results, recommendations are outlined for delimiting the relevant population for FSC casework involving West Yorkshire speakers. With regards to speech accommodation in FACE and /t/, results show that accommodation behaviour is highly variable across participants, both in terms of the direction and amount of accommodation present. All participants were considered to accommodate in at least one speech parameter, and a small number of participants displayed very high levels of within-speaker variability across tasks, highlighting the level of potential impact that speech accommodation can have on socially salient speech parameters. The consequences of these findings are addressed from both a FSC casework perspective and also in terms of sociolinguistic research practices more generally.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Δ	Change or difference
BDM	Bark Difference Metric
DID	Difference in distance
DyViS	Dynamic Variability in Speech
EDAC	English Dialects App Corpus
f_0	Fundamental frequency
F1-3	Formants (i.e. F1 = first formant)
FSC	Forensic speaker comparison
HCI	Human-to-computer interaction
HCRC	Human Communication Research Centre
HKE	Hong Kong English
LADO	Language analysis for the determination of origin
MFCC	Mel-frequency cepstral coefficient
RP	Received Pronunciation
SD	Standard deviation
SDS	Spoken dialogue systems
SSBE	Standard Southern British English
VOT	Voice onset time
WYRED	West Yorkshire Regional English Database

Acknowledgements

Firstly, I would like to express my gratitude to my supervisor, Dr Erica Gold, for all of her support and guidance over the years. From working with her as a master's student in York, to being part of the WYRED project and becoming her PhD student, she has always believed in me and helped me to pursue my goals. Thank you for all of the time you have taken to read and edit drafts of this thesis and for every supervision meeting where new ideas were discussed and developed. I feel very lucky to have such an inspiring mentor and friend.

I would also like to thank Sula Ross for all of her advice and encouragement. Working alongside you and Erica on the WYRED project whilst completing my PhD helped to keep me motivated and meant that the PhD process was never a lonely one. Related to this, I am indebted to all of the participants from West Yorkshire who volunteered their time to take part in the WYRED project. I would also like to thank everyone who assisted with recruitment and transcription of the recordings and special thanks to Lucy Jackson for transcribing all of the researcher files that I analysed in this project.

I would like to thank the whole Linguistics team at the University of Huddersfield for providing a friendly and collaborative environment in which to work and study. I am also extremely appreciative of the many opportunities I was afforded in my research assistant role and as a PhD student, including the chance to travel around the world to attend conferences and workshops, and to teach on the Sociolinguistics module. I would also like to thank everyone at J P French Associates for allowing me the flexibility to work part-time whilst writing this thesis. Since being welcomed to the team, I have had many useful discussions about my research and I am very appreciative of all the words of advice and feedback that I have received.

There are plenty of other people who have provided guidance over the course of my time studying for this PhD. I am grateful to Dr Sam Kirkham and Dr Danielle Turton who both gave me valuable advice relating to the design of my study and the statistical methods applied. Dr Lauren Ackerman also provided helpful statistical advice. I am also thankful for the interesting discussions that I have had with Dr Grace Cao about speech accommodation research, and with Dr Christin Kirchhübel about my project as a whole.

I must also thank all of my friends and family for their love and encouragement over the last few years. Anna, Nina and Kat have been my biggest cheerleaders and have always been there whenever I needed a break from it all. Thanks also to Kati for all of the motivational messages and reassurance. Thank you to Mum, Dad and Emily for everything. If it weren’t for your support I would never have completed an MSc in forensic speech science and discovered my love for forensic phonetics. Last but definitely not least, thank you to Adam for all of your love and kindness over the years. You have always been by my side (even before the lockdowns meant that you had no choice!) and I will forever be grateful for your support.

Declaration

This is to certify that this thesis contains original work and that all contributions from external sources are referenced appropriately. I also declare that aspects of the research have been previously published or submitted to journals and conference proceedings. These publications are as follows:

Earnshaw, K. (forthcoming). Examining the implications of speech accommodation for forensic speaker comparison casework : A case study of the West Yorkshire FACE vowel. *Journal of Phonetics*.

Earnshaw, K., & Gold, E. (2019). “We don’t pronounce our T’s around here": Realisations of /t/ in West Yorkshire English. In *Proceedings of ICPhS 19, Melbourne, Australia* (pp. 2655–2659).

Earnshaw, K., & Gold, E. (2018). Variation in the FACE vowel across West Yorkshire: Implications for forensic speaker comparisons. In *Proceedings of Interspeech 2018, Hyderabad, India* (pp. 2743–2747).

1. Introduction

This thesis is the first of its kind to explore speech accommodation in forensically-relevant speaking tasks that differ not only by interlocutor but also by context of the interaction and speaking style. In addition, this research examines regional variability across three neighbouring speech communities in West Yorkshire with respect to two parameters that are believed to be salient in this region: the FACE vowel and intervocalic /t/. The purpose of this research is to consider how much these two speech parameters vary between speakers from different boroughs of West Yorkshire, and within speakers across different speaking situations. While this research is intended to contribute to the sociophonetic literature in relation to speech accommodation and West Yorkshire English, one of the main motivations for this research is to investigate the implications of speech accommodation for forensic speaker comparison (FSC) casework. The extent to which regional differences exist at a local level is also evaluated from a FSC perspective.

Section 1.1 of this introduction chapter provides an overview of the FSC task and the potential challenges that speech accommodation may bring. Section 1.2 sets out the overarching aims and research questions addressed in this thesis and the structure of this thesis is outlined in Section 1.3.

1.1. Forensic speaker comparison

FSC casework usually involves the analysis of two or more speech samples, including a reference sample of a known suspect and a questioned sample containing an unknown, alleged offender's speech. As part of this analysis, the expert must assess the probability of obtaining the speech evidence under the hypothesis that the samples came from the same speaker versus the hypothesis that they came from different speakers. The purpose of this task is to form an expert opinion which can be presented in court to assist the triers of fact in deciding whether or not the suspect is the person heard in the questioned recording. For a comprehensive account of the history of FSC in the UK and its developments from the 1960s to present day, please see French (2017).

A range of different methods of analysis can be used for FSCs, with the combined auditory and acoustic phonetic approach being the most prevalent across the world and the only approach used in the UK for evidential purposes (Gold & French, 2011; Morrison et al., 2016). Using this combined approach, the expert examines a wide range of phonetic and non-phonetic features. These can include segmental features (vowels and consonants), suprasegmental features (fundamental frequency, voice quality, intonation and tempo), higher order linguistic features (use of discourse markers, telephone opening and closing behaviours, lexical usage) and non-linguistic features (filled pauses and other disfluency features). For further details relating to the specific methods of analysis that are often undertaken for each of these different features, see Gold & French (2011, pp. 301–302). Also see Foulkes & French (2012) for further information about the use of the linguistic-acoustic method in FSC casework. The specific features that are analysed in any given case will be determined by the expert, depending on the quality and quantity of speech available in the evidential samples, and therefore the range of features examined tends to vary on a case-by-case basis.

On the basis of their analyses, the expert will arrive at an accent profile for the speaker in each of the evidential recordings before assessing the similarities and differences between the samples. In order to evaluate the strength of the evidence, the expert must also consider how typical any shared features are in relation to the wider relevant population, based on the accent profile that has been identified for the questioned speaker. If the shared features are considered to be unusual in the relevant population, this would provide stronger evidence in support of the same speaker view than if the shared features were relatively typical for most speakers in this population. Typicality can be evaluated on the basis of an expert's knowledge and experience and/or with reference to population data from relevant, high quality data sets (where these exist) (c.f. Codes of Practice (Forensic Science Regulator, 2021)). When considering typicality, it is crucial that an appropriate reference population is selected in order to avoid under- or overestimations of the strength of evidence. Accordingly, an important area of research in this field is to understand the extent to which local level regional variation exists, in order to inform how narrowly or broadly reference populations need to be defined. Knowledge of regional variability is also vitally important for making an assessment of a speaker's accent profile.

Before undertaking a FSC, the expert must assess the adequacy of the samples taking into account any inconsistencies such as differences in technical quality, channel mismatches (telephone recording versus direct recording), as well as differences in speaking styles and speech modes (normal speaking level versus elevated voice, shouting or whisper etc.). In cases involving extreme mismatches between samples, where few to no features can be reliably compared, the samples are deemed unsuitable for analysis. However, in many instances it is possible to compare samples involving inconsistencies of this nature, so long as the expert considers the potential consequences of these discrepancies when interpreting the findings. In cases such as these, experts can refer to previous studies which have documented within-speaker variability in a range of different scenarios, in order to evaluate whether any differences observed indicate that different speakers are involved or whether they can instead be adequately accounted for by within-speaker variability.

Stylistic variation is an important source of within-speaker variability that often needs to be accounted for in FSC casework. Empirical quantitative sociolinguistic studies of stylistic variation initially tended to focus on how speech varied across situations involving different levels of "attention paid to speech" (Labov, 1966, 1972). Under this framework, Labov made the distinction between *casual speech* which he defined as "the everyday speech used in informal situations, where no attention is directed to language" (1966, p. 100) versus *careful speech* which he described as "the type of speech which normally occurs when the subject is answering questions which are formally recognized as 'part of the interview'" (1966, p. 92). However, many scholars found that in practice it was often difficult to reliably categorise their data into one of these groups. In the late 1970s and 1980s, the focus of stylistic variation shifted from placing styles on a single dimension, measured by the amount of attention paid to speech, to style being primarily influenced by the addressee. This conceptualisation of style as audience design was proposed by Bell (1984). Under this framework, it was asserted that speakers style-shift primarily in response to their audience, adjusting their speech style not only in response to their addressee but also in response to what Bell termed "auditors, overhearers and eavesdroppers" (1984, p. 159).

Bell's (1984) audience design model had strong foundations in Speech Accommodation Theory (SAT) which was proposed by Giles and colleagues in the late 1970s (Giles, 1973; Giles, Mulac, Bradac, & Johnson, 1987; Giles, Taylor, & Bourhis, 1973; Giles & Powesland, 1975). SAT focusses on the social cognitive processes motivating speech accommodation, whereby speakers adjust their speech according to whom they are talking. This phenomenon has been studied by sociolinguists, psycholinguists and cognitive scientists and research has found that accommodation can occur at various language levels including lexical choices, syntactic structure, prosodic features and phonetic parameters. Unfortunately, however, the fields of forensic speech science and speech accommodation research have rarely intersected. Although it is generally recognised by forensic experts that differences in speaking styles need to be taken into account in FSC casework, no studies to date have specifically examined stylistic variation across forensically-relevant contexts with the emphasis being placed on the influence of the interlocutor. Therefore there is limited forensically-relevant research available which relates to how speakers might be affected by speech accommodation in the kinds of scenarios that are typically encountered in FSC casework.

There are a number of reasons why it is difficult to apply the findings of many traditional accommodation experiments to FSC casework. Numerous accommodation studies focussing on the direct influence of the interlocutor involve highly controlled laboratory tasks which do not accurately reflect real-life scenarios faced in FSC casework (c.f. Babel & Bulatov, 2012; Cao, 2018; Levitan & Hirschberg, 2011; Pardo et al., 2018b; Weatherholtz, Campbell-Kibler, & Jaeger, 2014 *inter alia*). Although there are also a range of studies that examine accommodation in less controlled settings, these often tend to involve comparisons of variables across fairly consistent speaking situations which are not directly relevant to FSC casework. For example, Trudgill (1986) analysed his own speech across 10 sociolinguistic interviews that he conducted in Norwich. Rickford & McNair-Knox (1994) examined short-term accommodation in two sociolinguistic interviews of one speaker, undertaken by different interviewers. Similarly, Llamas, Watt, & Johnson (2009) examined the variable linguistic behaviour of five speakers in three interviews each conducted by separate interviewers. The speech data analysed in studies such as these does not adequately represent the range of contexts that often occur in FSC cases, where the interlocutor is not usually the only factor that has the potential to introduce variability across the evidential samples.

In the context of FSC casework in the UK, the majority of known and questioned samples submitted for analysis involve the speaker of interest talking to another person as part of an interactive conversation (J.P. French, personal communication, 2020). Furthermore, in most of these cases, the person that the known speaker is talking to is not normally the same person that the questioned speaker is talking to. A large proportion of UK FSC cases involve reference speech samples of the known speaker interacting with a police officer. At least two thirds of reference samples are typically from police interview recordings and additional sources of reference recordings include speech samples from police custody footage, police body-worn cameras and 999 telephone calls (J.P. French, personal communication, 2020). As a result of this, known and questioned material will generally tend to involve different interlocutors and the speaking styles and levels of formality will rarely be consistent across samples. For instance, in contexts involving police officers there will often be a question and answer format whereas questioned samples may be less structured. The question and answer format of police interviews can also introduce an imbalance in terms of power dynamics between interlocutors which can potentially lead to speech adaptations that need to be accounted for in FSC analyses. Other factors such as the emotional state of the speaker of interest or whether the interaction took place face-to-face can have further consequences for FSC casework.

FSC cases can also involve the analysis of recorded speech samples containing only one voice, such as answer messages or, increasingly more frequently, voice notes. A distinction can be made between these two forms of communication, with the more traditional answer message recording being left when a call recipient does not answer the phone, while voice notes are recordings that are made and sent directly via messaging apps such as WhatsApp, Facebook Messenger and Telegram. Voice notes are often considered to be a quick and easy way to communicate instantly, as well as being a helpful alternative to typing long text messages. Often, voice notes are sent between recipients forming an interactive, live conversation, whereas answer messages do not involve any instant feedback from another person. However, in some FSC cases it is not known whether a voice note has been sent as part of an interactive conversation or whether it is a standalone recording that is more comparable to an answer message. Nevertheless, in both of these types of speech samples there is usually an intended recipient and therefore some influence of audience design may be at play, despite there not necessarily being a live interlocutor. As FSC cases can sometimes involve comparing interactive, live conversations with more monologue-style, non-interactive speech samples, it is important that the ways in which a person's speech may vary across these differing contexts are understood.

A greater understanding is required of how speech may change across scenarios such as those described above, where speakers are not only interacting with a different person but other factors such as the speaking style, speaking mode and channel may also vary between evidential samples. Extreme forms of speech accommodation either towards or away from an interlocutor over the course of an interaction could result in high levels of variability within a sample, making it difficult for an expert to arrive at a reliable accent profile for the speaker. Furthermore, in cases where only a limited range of features can be reliably compared, accommodation behaviour could lead to evidence in support of the different speaker view (or at least reduce the strength of the evidence in support of the same speaker view), despite the samples being produced by the same speaker. In much the same way that cases involving attempted disguise or imitation can pose a challenge for the expert, the consequences of speech accommodation also have the potential to lead to the expert either only being able to reach a limited conclusion or resulting in an inconclusive analysis. For this reason, further research into speech accommodation using more naturalistic speech data is required in order to understand how much certain speech parameters could change across different forensic samples. The more we know about the effects of accommodation and stylistic variation, the better our understanding will be of the implications this could have on cases involving mismatched samples.

1.2. Research aims

The first aim of this research is to collect population data for West Yorkshire, specifically in relation to the FACE vowel and intervocalic /t/, and to establish how generalisable this data is across the West Yorkshire boroughs of Bradford, Kirklees and Wakefield. This is intended to be of use for forensic caseworkers as it will provide an indication of whether the three boroughs need to be treated as separate populations (for establishing typicality) or if they can instead be grouped together as "West Yorkshire English" when delimiting the relevant population for FSC casework involving West Yorkshire speakers. Although it is not possible to define a clear threshold for delimiting the three areas, in terms of how many features would need to be consistent or variable across boroughs, research has indicated that FACE and intervocalic /t/ were previously some of the most regionally-stratified features across West Yorkshire (Petyt, 1985) and therefore if they do not appear to vary across the three boroughs, this would suggest that the boroughs could be grouped together.

It should be acknowledged that FSC conclusions are reached on the basis of a holistic assessment of similarity and typicality across a whole constellation of features rather than any one individual feature. However, it is important to note that both FACE and intervocalic /t/ are features that would usually be examined as standard, as part of a segmental analysis in a FSC case. Further motivations for analysing these specific features are outlined in Chapter 3. It is anticipated that the findings of this investigation will be of use not only to forensic speech scientists but also to sociophoneticians more generally, as the extent to which speech production varies on a local level in West Yorkshire is currently under-researched. This investigation constitutes the first major study of West Yorkshire English for a generation and provides an update to previous descriptive accounts of the accents in this area (c.f. Broadbent, 2008; Petyt, 1985; Watt & Tillotson, 2001).

The second aim of this research is to establish how much within-speaker variability is present across a range of speaking tasks designed to reflect forensically-relevant speaking situations which involve different contexts, interlocutors and channels, as well as varying levels of formality, pressure on the speaker and differing roles in the conversation. The purpose of this is to explore how short-term accommodation behaviour might manifest itself in phonetic productions within contexts commonly encountered in FSC casework and to highlight where caution may be required. In line with most empirical research involving semi-spontaneous speech, it is not possible to disentangle the effects of the interlocutor and speech style from one another. Although these two factors are somewhat independent (i.e. a speaker can vary stylistically with the same interlocutor and also be consistent stylistically with multiple interlocutors), it is not always apparent which factor can best account for any within-speaker variability. Bell's (1984) model of audience design suggests that the primary factor in a speaker's style-shifting is the addressee, however, "nonpersonal" factors such as topic, setting and degree of attention are also recognised as having an effect. For this reason, in this study the term “accommodation” is defined broadly to refer to changes that occur across multiple speaking situations involving different interlocutors. The observations underpinning this investigation are subsequently evaluated in order to consider the extent to which speech accommodation could impact FSC casework.

Four main research questions will be addressed in this thesis, with the first two relating to the first research aim and the third and fourth questions relating to the second research aim. These research questions are as follows:

1. How much local level variation exists across the West Yorkshire boroughs of Bradford, Kirklees and Wakefield with respect to the FACE vowel and intervocalic /t/?
2. Should reference population data for West Yorkshire speakers be separated by metropolitan borough or is it appropriate to group the boroughs more broadly as “West Yorkshire English”?
3. To what extent do speakers adapt their FACE and intervocalic /t/ productions across forensically-relevant scenarios involving different interlocutors and speaking styles?
4. What are the potential implications of speech accommodation for FSC casework?

1.3. Thesis structure

This thesis is made up of nine chapters. Chapter 2 provides further context for this research by providing a detailed review of the speech accommodation literature. The chapter describes previous speech accommodation studies and highlights gaps in the literature which are addressed in this thesis. Chapter 3 gives details of the data used for all of the investigations presented in this thesis. Information is provided about the area of West Yorkshire, the fieldwork that was conducted to collect the West Yorkshire Regional English Database (WYRED), as well as details relating to the specific WYRED participants analysed in this thesis and the speaking tasks that they each completed. Details are also provided about the analytical approach applied in the case studies presented in this thesis. The four subsequent chapters involve individual case studies relating to regional variation and speech accommodation in the FACE vowel and intervocalic /t/, respectively.

Chapter 4 establishes how FACE is realised in West Yorkshire and explores how much local level variation is present across the region. The motivations for investigating this parameter are explained and the findings of previous studies which have examined FACE are summarised, while highlighting regional variation across the UK as well as social and linguistic factors that condition this variable. These factors are all taken into account in this analysis and a range of quantitative methods are applied to determine how the FACE vowel varies across the region. The implications of the results of this case study are discussed before moving onto the analysis of phonetic accommodation in FACE. Chapter 5 explores the within-speaker variability in FACE productions across a range of tasks and considers how the participants' realisations relate to those of their respective interlocutors. The influence of exposure to the interlocutor is evaluated by considering how consistent FACE realisations are within tasks, and how FACE varies across tasks and from the first half of a task to the second half. The results of this investigation are presented at the group level and a selection of individual results are also discussed.

Chapter 6 investigates how intervocalic /t/ is realised across the West Yorkshire boroughs of Bradford, Kirklees and Wakefield whilst also considering how the phonetic environment influences productions. Previous findings in relation to regional, social and linguistic variation in /t/ are outlined and the motivations for examining this variable are provided. The methods of analysis are described and the results of this investigation are presented. The implications of the findings for the fields of forensic speech science and sociophonetics are also described. Chapter 7 builds on the findings of Chapter 6 by exploring the extent to which intervocalic /t/ productions vary across tasks in order to establish the effects of the interlocutor and the speaking style on this phonetic variable. Accounting for the influence of phonetic environment, within-speaker variability is explored within and across tasks, at the group level and at the individual level. Again, the findings of this investigation are discussed in relation to how intervocalic /t/ productions could be influenced by speech accommodation in the context of FSC casework.

Chapter 8 draws together the findings from Chapters 4-7 in order to address each of the main research questions posed in this thesis. Findings in relation to levels of regional variability and speech accommodation are compared across the two phonetic variables examined in this investigation. It is then considered how generalisable the population data for FACE and intervocalic /t/ are across the West Yorkshire boroughs. Consistency within the participants' accommodation behaviour is examined across tasks for FACE and intervocalic /t/, and across these two speech parameters. A fictional FSC case study is also presented by analysing the speech of one specific individual who displayed particularly high levels of accommodation overall. The potential implications of speech accommodation for FSC casework are subsequently reflected upon. The limitations of the research presented in this thesis are acknowledged and ideas for further research are set out. Finally, Chapter 9 outlines the conclusions of this thesis. It summarises the scope of the research and highlights the key findings. This chapter also describes the contributions that this research makes to the fields of forensic speech science and sociophonetics.

2. Literature review

The aim of this chapter is to situate the research presented in this thesis within the wider context of previous areas of study relating to variationist sociolinguistics and speech accommodation, and to highlight gaps in the literature that are addressed in this investigation. Section 2.1 provides an in-depth review of the speech accommodation literature, detailing what the phenomenon is and how it has been studied previously. Section 2.2 presents an explanation of the potential practical implications that speech accommodation may have in the fields of forensic speech science and sociophonetic research more broadly as well as briefly discussing implications for language analysis for the determination of origin and the development of spoken dialogue systems.

2.1. Speech accommodation

As mentioned in the introduction chapter, speech accommodation is a process whereby speakers adapt their speech production according to whom to they are talking (Giles, 1973; Giles & Powesland, 1975; Trudgill, 1981). In broad terms, speech accommodation can be can lead to two main outcomes: *convergence* and *divergence*. Convergence is a form of variation in speech production whereby a talker adopts aspects of their audience or interlocutor's speech, resulting in a reduction in linguistic/phonetic distance between the speakers. Divergence, on the other hand, is a form of speech accommodation where the linguistic/phonetic distance between speakers is increased as a result of a talker producing speech in a way that is dissimilar to their interlocutor. The term *maintenance* is also sometimes used in accommodation research to refer to when a speaker neither converges nor diverges but instead remains consistent in their speaking style.

It should be noted that in some research the term *accommodation* is used to refer specifically to speech adaptation towards an interlocutor; i.e. convergence. In this thesis, the term *accommodation* is defined more broadly as a general adaption of speech which can lead to interlocutors becoming more or less similar to one another. The terms *convergence* and *divergence* are used when discussing the specific direction in which the speaker has shifted. Furthermore, in this thesis a similar approach to Cao (2018) is adopted, whereby the terms *convergence* and *divergence* are not intended to express an opinion regarding the motivations of the speaker but instead they simply reflect the fact that a change has been observed.

It is necessary to make a distinction here between short-term and long-term speech accommodation. As the labels suggest, short-term accommodation refers to specific changes that occur in real time over the course of an interaction (whether that be in the form of a conversation or a non-interactive speaking task); whereas long-term accommodation refers to the persistence and maintenance of a specific change over a longer period of time after exposure to an interlocutor/group of interlocutors. To put this another way, Auer & Hinskens (2005) state that "short-term accommodation becomes long-term accommodation as soon as it permanently affects the accommodating speakers." (2005, p. 335).

Although it is fair to say that the majority of experimental accommodation research has focussed on short-term accommodation, where changes are observed on one particular occasion; both types of accommodation have been studied empirically. For instance, Evans & Iverson (2007) observed long-term phonetic accommodation taking place amongst a group of students from the Midlands who had moved away to university and began to interact with speakers of Standard Southern British English (SSBE) on a regular basis. Over a period of two years, it was observed that the speakers began to converge towards the SSBE speakers in terms of their vowel productions. Pardo, Gibbons, Suppes, & Krauss (2012) similarly investigated accommodation amongst college roommates over a period of approximately six months. Interestingly, in this study it was found that phonetic convergence in vowel productions correlated with roommates' self-reported closeness to a moderate degree, with perceptual detection of convergence varying across different linguistic items.

Long-term accommodation has also been investigated in the context of second language acquisition (Chang, 2012; Nguyen & Delvaux, 2015; Sancier & Fowler, 1997) and in terms of whether it can be interpreted as a driving force for language change over time (Auer & Hinskens, 2005; Babel, McGuire, Walters, & Nicholls, 2014; Delvaux & Soquet, 2007; Tamminga, 2016; Trudgill, 1986). The findings of these studies and potential implications of long-term accommodation will be briefly discussed in Section 2.2.1. However, it is worth noting here that this thesis is primarily interested in investigating short-term accommodation to other speakers within an interaction and therefore a strong emphasis is placed on short-term accommodation studies throughout this chapter.

As this thesis is interested in the consequences of short-term speech accommodation in forensic contexts, it is perhaps also useful to draw some clear distinctions here between speech accommodation and speech disguise. Firstly, it should be noted that the term *speech accommodation* can be used to refer to shifts in speech that occur either consciously or subconsciously when interacting with another speaker; whereas the term *speech disguise* refers to someone deliberately attempting to conceal their identity in order to avoid detection. Secondly, speech accommodation involves within-speaker changes which are exclusively created by the speaker (e.g. articulatory movements leading to changes in pitch, voice quality and accent features); whereas disguise can either be achieved internally by the speaker or externally by using objects to distort the speech, such as face-concealing garments or electronic devices. Although it is possible that similar phonetic outcomes may be achieved, it is expected that imitation for the purposes of disguise will generally be less subtle than speech accommodation. For instance, speakers attempting to imitate another speaker by disguising their speech will often only adopt the most stereotypical and socially salient features identified for the individual being impersonated, depending on the degree of their phonetic talent, whereas convergence will often tend to be much less obvious so as not to offend the interlocutor. The focus of the present study is to investigate short-term phonetic accommodation in contexts where accommodation might be expected to occur naturally, without a speaker directly attempting to imitate their interlocutor or disguise their speech.

2.1.1. Why speech accommodation occurs

Since the social psycholinguist Howard Giles and his colleagues released their initial publications concerning Speech Accommodation Theory (SAT; Giles, 1973; Giles, Mulac, Bradac, & Johnson, 1987; Giles, Taylor, & Bourhis, 1973; Giles & Powesland, 1975) there has been wide debate regarding the underlying mechanisms that cause accommodation behaviour. While some scholars argue that accommodation is an automatic phenomenon which occurs without conscious control; others maintain that it is socially-motivated and is therefore to some degree controlled, whether consciously or unconsciously. In this context, “conscious” accommodation behaviour refers to a speaker deliberately adapting their speech depending on their interlocutor in order to achieve social goals, whereas “unconscious” accommodation refers to shifts that take place without the speaker specifically intending to alter their speech. Much of the research in this area has been devoted to examining convergence, which is sometimes also referred to as *entrainment, imitation, priming, coordination* or *alignment,* across different research disciplines (c.f. Finlayson, Lickley, & Corley, 2012; Gijssels, Casasanto, Jasmin, Hagoort, & Casasanto, 2016; Nenkova, Gravano, & Hirschberg, 2008; Pickering & Garrod, 2004).

According to Giles et al. (1987), SAT focusses on the social cognitive processes motivating speech shifts during social interactions and the social consequences of these (1987, p. 14). Within SAT, an account of convergence was outlined which suggested that speech shifts towards interlocutors during social interactions are the consequence of socially-motivated strategies for “evoking listeners' social approval, attaining communicational efficiency between interactants, and maintaining positive social identities.” (1987, p. 15). It was further claimed that “it is the individual's perception of the other's speech that will determine his or her evaluative and communicative responses.”(1987, p. 15). Giles, Coupland, & Coupland (1991) also proposed that if a speaker wanted to disassociate themselves from their interlocutor they could take the opposite approach and instead diverge away from their interlocutor (1991, p. 2).

One of the initial publications related to SAT was Giles’ (1973) paper which described the Accent Mobility model. This paper conveyed the idea that accent mobility (i.e. the ability of an individual to modify their accent or pronunciation) was person-based as well as being context-based. Specifically, this model stated that most individuals’ accent mobile behaviour shifts along a “standard-broad regional pronunciation continuum and also reflects social prestige values” (Giles, 1973, p. 90). Related to this, was Bell's (1984) Audience Design model which suggested that stylistic-variability could be accounted for by speakers adjusting their speech style not only in response to their direct addressee but also to any other individuals assuming the role of auditor, overhearer or eavesdropper.

In 1987, Giles et al. proposed that SAT should be relabelled as Communication Accommodation Theory (CAT), in order to more adequately reflect development of the theory since Giles' Accent Mobility model and the expansion of research in this area to include non-verbal accommodative behaviour. In 1991, Giles et al. highlighted how wide-spread convergence had been shown to be by providing a list of features that had converged in various studies, including non-verbal features such as gesture (Maurer & Tindall, 1983), head nodding and facial affect (Hale & Burgoon, 1984) and posture (Condon & Ogston, 1967). Prior to this, Trudgill (1986) had also noted that scholars in fields such as communications and psychology had investigated accommodation behaviour in other non-linguistic features such as body movement, proximity, gaze direction and eye contact and he argued that research such as this demonstrated that convergence of this type is a "universal characteristic of human behaviour" (1986, p. 2).

In contrast to theories such as SAT (Giles, 1973) and CAT (Giles, Coupland, & Coupland, 1991; Shepard, Giles, & LePoire, 2001), which focus on the role of social identity and cooperation to achieve social goals, cognitive scientists and psycholinguists have proposed that convergence is a consequence of automatic cognitive mechanisms that function to enable language processing and communication. For instance, Pickering & Garrod (2004) developed a psycholinguistic account of interactive language processing known as the "interactive-alignment model" which hypothesises that successful communication in dialogue involves the alignment of interlocutors' representations, via priming at numerous levels of linguistic representations. This theory suggests that alignment takes place automatically and that relatively few conscious or deliberate strategies are required. More specifically, they propose that the automaticity of alignment may take place at "the post-conscious level, whereby automaticity requires awareness of the stimulus when it originally occurred" meaning that "interlocutors have to attend to what the other is saying in order for automatic alignment to occur" (Pickering & Garrod, 2004, p. 214). Following on from this, Pickering & Garrod (2004) state that this post-conscious notion of automaticity can explain why alignment is expected to be affected by interlocutor-specific factors as well as by social factors, even when the interlocutors are not aware that they are aligning (2004, p. 214).

A range of other theoretical accounts have been called upon to explain how and why speech accommodation takes place from a cognitive perspective. These include the Episodic Theory of speech perception and production (Goldinger, 1998), the Motor Theory of speech perception (Liberman & Mattingly, 1985) and the Direct Realist Theory (Fowler, 1986). In 2013, Gambi & Pickering advocated the need to integrate low-level, mechanistic accounts (e.g. Goldinger, 1998; Liberman & Mattingly, 1985) with higher-level accounts such as CAT (Giles et al., 1991; Shepard et al., 2001) in order to more adequately explain speech convergence. Subsequently, Gambi & Pickering (2013) recommended an integrated theory of production and comprehension, known as the Simulation Theory of speech perception which was formulated by Pickering & Garrod (2013). According to the this theory, "perception of other people's speech involves covert simulation of their speech, and covert simulation is achieved by running forward models of one's own speech production system." (Gambi & Pickering, 2013, p. 4). A crucial aspect of this theory is that language production and comprehension are inextricably linked as a result of the processes required in self-monitoring of speech production. Although it is not possible to provide a comprehensive account of this model here, the general principle is that alignment occurs as the result of an automatic, cognitive reflex which encourages similarity between interlocutors across all linguistic representations.

Empirical speech accommodation studies have demonstrated that convergence can take place in asocial settings where there is no physical interlocutor present including speech-shadowing tasks (Goldinger, 1998; Namy, Nygaard, & Sauerteig, 2002; Shockley, Sabadini, & Fowler, 2004) and in experiments involving human-to-computer interaction where participants responded to a virtual interlocutor (Cowan & Branigan, 2015; Gessinger, Mobius, Fakhar, Raveh, & Steiner, 2019; Staum Casasanto, Jasmin, & Casasanto, 2010). These findings would appear to contradict the idea that convergence is an entirely socially-motivated process which is adopted by a speaker as a strategy for trying to build rapport or to coordinate with an interlocutor. It would also perhaps provide evidence to support the idea that convergence may be, at least in part, an automatic process. However, in the following section of this chapter a number of studies are described which provide evidence to suggest that accommodation behaviour is to some extent mediated by a wide range of social factors related to both the individual and their knowledge about and attitude towards their interlocutor. For this reason, it is considered that speech accommodation is both internally and externally motivated.

For further discussion of the history of the theory underpinning the study of speech accommodation research and methodological approaches applied in recent studies, see Solanki (2017; secs. 2.1 & 2.2 respectively) which provides a thorough account of these topics. Cao (2018, Chapter 4) also presents a systematic review of theories related to short-term accommodation.

2.1.2. How speech accommodation is studied

This section sets out a range of different experimental approaches that have been applied to conduct speech accommodation research and describes the different types of speech data that have been examined. The aim of this section is to exemplify how speech accommodation studies have previously been conducted and how the research presented in the current thesis links to existing studies and as well as to emphasise the ways in which this research is unique.

2.1.2.1 Data

Convergence has been investigated extensively using speech elicited in non-interactive laboratory tasks and it has also been examined in conversational interaction to a lesser degree (Pardo et al., 2018b). This sub-section describes a range of methods that have been used to collect non-interactive speech data and conversational data for the purposes of examining speech accommodation. In outlining these methods, a number of previous studies from sociolinguistic and psycholinguistic perspectives are referenced and described, with an emphasis on short-term speech accommodation research, as this is the focus of the present thesis. However, reference is also made to some long-term accommodation studies. The advantages and disadvantages of using non-interactive and conversational data are also addressed.

2.1.2.1.1 Non-interactive

Non-interactive speech elicitation often involves speech-shadowing tasks, which are also sometimes referred to as auditory naming tasks. In these tasks, participants typically produce *pre-exposure baseline utterances*, by reading aloud a list of printed words or sentences, before subsequently producing *shadowed utterances*, prompted by recordings of utterances spoken by a model talker. The baseline and shadowed utterances of each target item can then be compared to see whether the participant converged towards the model talker or not. If the shadowed utterances are determined to be more similar to those of the model talker than the baseline utterances are, it can be concluded that convergence took place. In some cases, participants also produce *post-exposure utterances,* by reading aloud the original list of printed words or sentences after the speech shadowing has taken place. The purpose of this is to enable the researcher to examine whether any changes from the baseline utterances to the shadowed speech are maintained afterwards. Goldinger (1998) was one of the first studies to investigate speech imitation by employing this speech-shadowing elicitation technique. In this study, it was found that low-frequency words elicited greater phonetic convergence than high-frequency words, and this finding was later replicated in a follow-up study by Goldinger & Azuma (2004). It was argued that this evidence provided support for the idea that memory preserves detailed traces of spoken words.

Since Goldinger's seminal study, a vast number of psycholinguistic studies have examined speech accommodation using data recorded as part of a speech-shadowing task (Babel, 2010, 2012; Babel & Bulatov, 2012; Babel et al., 2014; Goldinger & Azuma, 2004; Namy et al., 2002; Nielsen, 2007, 2011; Shockley et al., 2004; Yu, Abrego-Collier, & Sonderegger, 2013). While the broad set-up described above has been applied in the majority of non-interactive accommodation studies, the precise nature of the experimental design in terms of the number of model talkers, number of shadowers and types of lexical items presented has varied greatly between studies. For a detailed summary of non-interactive speech-shadowing studies of phonetic convergence, see Pardo, Urmanche, Wilman, & Wiener (2017) which consolidates the findings of almost thirty accommodation studies in a large-scale examination of the impacts of talker sex, word frequency, and model talkers on multiple measures of convergence.

Most speech-shadowing accommodation studies are entirely asocial, in that the participants can only hear the speech of the model talker and cannot see them. This enables the researcher to ensure that the participants only respond to the model talker's speech, as opposed to any other form of input. However, Babel (2012) adapted the original speech-shadowing paradigm by presenting a subset of her participants with an image of a model talker, in order to investigate whether this has any impact on the degree to which they converged towards the model talker. In this study, participants were assigned to either a black model talker or a white model talker and, within this talker manipulation, participants were divided into two separate conditions; one where participants were presented with a digital image of their assigned model talker and one without a visual prompt. Those participants assigned to the visual prompt condition were asked to rate the attractiveness of the model talker, after the speech-shadowing task was completed. Babel found that more convergence occurred in the visual prompt condition than in the condition without a visual prompt and that levels of convergence were subtly affected by attractiveness ratings. It was argued that these findings could be interpreted as providing evidence in support of the idea that speech accommodation is a socially driven phenomenon (Babel, 2012, p. 186).

Speech accommodation studies have also been carried out using non-interactive speech data collected without the use of a speech-shadowing task. For example, two investigations of long-term speech accommodation involved the analysis of read speech elicited by way of recording participants reading a set of experimental words and short reading passages, across four separate recording sessions (Evans & Iverson, 2007; Pardo et al., 2012). In the case of Evans & Iverson (2007), changes in accentedness ratings and vowel formants were observed amongst a group of students from the Midlands who had moved away to university. It was concluded that the participants' vowel productions typically became more similar to those of SSBE speakers, over a period of two years. In Pardo et al. (2012), phonetic convergence in five pairs of previously unacquainted college roommates was examined over a six-month period. Based on perceptual similarity ratings, it was observed that most roommates converged in perceived phonetic form during the second recording session, which took place after approximately six weeks of cohabitation, and all roommates converged by the final recording session. In both of these studies, the participants were perceived to have converged towards their peers over time, based on their read speech productions. In addition to laboratory-based recording settings, non-interactive speech data has also been elicited using online platforms.

Weatherholtz, Campbell-Kibler, & Jaeger (2014) employed a novel web-based paradigm to investigate syntactic alignment in a study of over 300 speakers. In this investigation, participants were initially presented with a spoken passage read aloud by one of three model speakers, in what was referred to as the *exposure phase*. Participants were subsequently required to complete a picture description task, in what was referred to as the *response phase*. During the response phase, ten line drawings were presented with four being designed to elicit a dative structure and the other six serving as fillers. Participants had to provide verbal descriptions of each of the ten drawings and these were recorded via the participant's web browser. The spoken productions were then examined to assess the effect of the syntactic structure of a passage participants heard during the exposure phase on their syntactic behaviour during the response phase. Participants were also required to complete a comprehensive social questionnaire to enable the researchers to test a number of social dimensions as predictors of alignment. Additionally, a selection of participants were assigned to a baseline condition where they completed the picture description task without any prior exposure to a spoken passage. This enabled the researchers to establish base rates of syntactic usage for each target picture so that these rates could be compared against those of the participants in the other experimental conditions. The results of this investigation revealed that there was an overall alignment effect across social conditions but the degree of alignment appeared to be mediated by a number of social factors.

The main advantage of the types of experimental design described above is that by eliciting speech in non-interactive settings, the researcher is able to carefully control the input that the participants receive and ensure that the participants produce the specific target sounds, structures or items under investigation. In cases where participants are presented with auditory stimulus, the use of pre-recorded speech spoken by a model talker enables the researcher to manipulate the experimental input to a high degree and to make sure that all participants are presented with the exact same input. However, it could be argued that speech elicited in non-interactive settings is less representative of the kind of speech that we might expect to find in spontaneous, everyday conversations. Furthermore, it is unlikely to be representative of the speaking styles that generally occur in evidential samples in FSC casework. While non-interactive speech data might be regarded as ideal for studies with a psycholinguistic focus, this type of data may be considered insufficient for investigations into the social motivations for accommodative behaviour, or for exploring the implications of accommodation for FSC casework. For this reason, a number of studies have examined accommodation using speech elicited in interactive conversations.

2.1.2.1.2 Interactive conversation

The majority of conversational data that has been analysed in previous accommodation studies has been recorded as part of an interactive paired task or game. Some accommodation studies have also made use of (semi-)spontaneous speech elicited in less structured conditions where speakers have been recorded having a casual conversation, for example. In some cases, the speech is recorded specifically for the purpose of assessing levels of accommodation and therefore participants are required to produce pre-exposure baseline utterances, shadowed utterances and post-exposure utterances. Other studies have made use of conversational data elicited for a purpose other than examining speech accommodation, including structured speaking tasks and spontaneous speech. In these cases, pre-exposure baseline utterances are not always available and therefore levels of accommodation are assessed in other ways.

The rest of this section is comprised of three sub-sections; the first two sections consider interactive conversational data which is elicited via structured speaking tasks, whereas the third section discusses how casual speech has been used to examine speech accommodation. The types of structured speaking tasks that have been used in previous accommodation studies can be separated into two different categories: human-to-computer interaction and human-to-human interaction. Both of these types of interaction are outlined below.

2.1.2.1.2.1 Collaborative speaking tasks

2.1.2.1.2.1.1 Human-to-computer interaction

Staum Casasanto et al. (2010) was one of the earliest studies to investigate speech accommodation within human-to-computer interaction (HCI). In this study, participants entered an immersive virtual reality environment and communicated with a virtual interlocutor named VIRTUO. Although participants engaged in an interactive conversation, they were made aware that their virtual interlocutor could not think, feel or process their accommodative behaviour. This data elicitation technique enabled the researchers to explore the extent to which interactional motivations drive accommodation by examining whether participants still accommodated when their behaviour could not be interpreted by their interlocutor. This study investigated accommodation in terms of speech rate by assigning participants to one of two conditions, where VIRTUO's speech rate had been digitally manipulated to be either fast or slow. From a practical perspective, this experimental design provided a high degree of control as the researchers could ensure that VIRTUO's speech rate was consistent across participants and that the interlocutor did not converge towards the participant; something which cannot be guaranteed in human-to-human communication. By comparing the participants' speech rate during baseline productions to when they were interacting with VIRTUO, it was observed that their speech rate was influenced by the speech rate of the virtual interlocutor. It was found that "participants who spoke to a fast-talking VIRTUO sped up significantly from their Baseline speech rate, and spoke significantly faster than their counterparts in the Slow condition" (Staum Casasanto et al., 2010, p. 130). One interpretation of this finding was that long-term social goals may be a factor that drives accommodation and therefore the participants may have been motivated to accommodate to VIRTUO by a general tendency to speak similarly to other speakers.

Other examples of accommodation studies which have employed virtual interlocutors include investigations of lexical alignment (Branigan, Pickering, Pearson, McLean, & Brown, 2011; Cowan et al., 2019; Cowan & Branigan, 2015), syntactic alignment (Cowan, Branigan, Obregón, Bugis, & Beale, 2015), alignment in fundamental frequency (f_0) (Gijssels et al., 2016; Ibrahim, Skantze, Stoll, & Dellwo, 2019) and convergence in allophonic contrasts and phenomena of local prosody (Gessinger et al., 2019). All of these studies found a tendency for participants to accommodate towards their virtual interlocutor, despite being aware that they were interacting with an automated system and therefore it is believed that there was no motivation to try to build rapport with the interlocutor.

Many HCI studies required participants to complete some form of game such as a picture naming and matching task, or a Guess Who game. However, Gessinger et al. (2019) employed a novel approach, where the experiment was presented to the participants as an application for learning the German language where participants were trained by a female called Mirabella. Participants were made to believe that the purpose of the task was to test the application, rather than for their speech to be analysed. This study used a Wizard-of-Oz paradigm to simulate an intelligent spoken dialogue system, so that participants believed they were interacting with an autonomous system; whereas in fact the experimenter actually had full control. The speech of Mirabella was all pre-recorded and available for the experimenter to play to the participant during the experiment. The participants were unable to see who they were talking to during the experiment as they only interacted with Mirabella's voice. The first two tasks elicited baseline productions of target items while the third and fourth tasks involved interactive tasks with Mirabella where the participant and Mirabella took turns to ask each other questions. The researchers were interested in a number of different phonetic features, however, Gessinger et al. (2019) only reported the results pertaining to the German [ɪç] vs. [ɪk] allophonic contrast. Overall, it was found that the two largest groups of participants showed either convergence to the spoken dialogue system with respect to the German [ɪç] vs [ɪk] contrast, or showed maintaining behaviour.

The main benefit of using virtual interlocutors to study accommodation behaviour is that this method of data collection is interactive and is therefore potentially more ecologically viable compared to a speech-shadowing or reading task; whilst nevertheless, the researcher is still able to control the input stimuli in the same way they can in a non-interactive task. The main disadvantage of this method is that the participant may not accommodate in the same way or to the same degree as they would if they were able to influence how their interlocutor thinks or feels about them, if we are to accept that accommodation is motivated to a large degree by social goals. Although it would appear that speakers do accommodate to virtual interlocutors, it is possible that social motivations to accommodate may be stronger when interacting with a human partner, in a face-to-face environment.

2.1.2.1.2.1.2 Human-to-human interaction

One of the most commonly used approaches to elicit shadowed utterances in live conversation has been to engage pairs of participants in a map-matching task by using modified versions of the Human Communication Research Centre (HCRC) Map Task (Anderson et al., 1991). The original HCRC map task involved pairs of participants each being given schematic maps, containing a range of overlapping landmarks, where one map had a route drawn on and one did not. It was the task of the participant without the route (the instruction receiver) to work together with their partner (the instruction giver) to determine the route from the start point, around various landmarks, to a finish point and draw it onto their map. Crucially, both participants were required to mention all of the landmarks in order to complete the task successfully. This is especially useful for analysing speech accommodation, as participants effectively complete a speech-shadowing task unconsciously, as part of a live conversation.

There are many examples of speech accommodation studies that have used a map-task technique to elicit data (Aguilar et al., 2016; Cao, 2015, 2018; Finlayson, Lickley, & Corley, 2012; Pardo et al., 2013a; Pardo et al., 2018b). In creating their own versions of the HCRC map task, researchers typically rename landmarks and place names in order to elicit the specific target sounds of interest in their investigation. For example, as part of an investigation into short-term accommodation in a range of different phonetic features, Cao (2018) examined participants' productions of the THOUGHT vowel by including landmarks in her map task such as the "Causeway Hotel", the "Thought Garden" and the "North Band". Some researchers have also made more extreme modifications to the HCRC map task in line with the specific aim of their study.

As part of a large-scale study of phonetic convergence in spoken communication, Pardo et al. (2018a) developed a role-neutral version of the HCRC map task, in order to create the Montclair Map Task Corpus. The main purpose of making the Montclair Map Task role-neutral was to balance the amount of speech that was elicited from each participant in the conversational pair, and to enable assessment of both participants' task performance. This move was motivated by an analysis of data from Pardo (2006), a study which used the HCRC map task, which found that instruction givers produced 2.7 times as many words as instruction receivers and twice as many landmark labels (Pardo et al., 2019, p. 379). Furthermore, both Pardo (2006) and a study by Pardo, Jay, & Krauss (2010) revealed an effect of conversational role whereby the participants in the instruction giver role tended to converge to a greater degree towards their interlocutor than the instruction receiver. Although this was an important finding regarding how accommodation behaviour can be affected by conversational role; the design of the Montclair Map Task aimed to remove this factor to facilitate future research into phonetic accommodation. In addition to this, it became possible to measure the task performance of both members of a pair using the Montclair Map Task, because both interlocutors have information to share and a task to complete. This consequently enabled researchers to examine how accommodative behaviour relates to communicative success.

Another example of a speech database containing data elicited via an adapted version of the HCRC map task is the Dynamic Variability in Speech Database (DyViS; Nolan, McDougall, de Jong, & Hudson, 2009). Task 1 of the DyViS database involved participants taking part in a simulated police interview where they assumed the role of the suspect being interviewed by one of two possible female researchers in the role of a police officer. Prompted by a series of visual aids in the form of maps and schemas intended to represent the suspect's memory and knowledge, the participants' task was to be as co-operative as possible in answering the officer's questions without revealing any incriminating information. In line with other map-task studies, this methodology enabled the researchers to elicit specific target phonetic variables during a socially rich, dyadic conversation.

One study which made use of the DyViS database to investigate accommodation behaviour from a forensic phonetic perspective was carried out by Gold, French, & Harrison (2013a), as part of a larger study into the speaker discriminant power of clicking behaviour for FSC casework. In this study, clicks were defined as a linguistic parameter as they were used as a discourse marker in conversation. Gold et al. (2013a) calculated click rates for 50 participants and their respective interlocutors across DyViS Tasks 1 and 2, in order to determine whether any increase in subjects' clicking behaviour might be accounted for by an interlocutor accommodation effect. Task 2 involved the participants having a telephone conversation with a male researcher in the role of an accomplice and the participant's task was to relay everything that had happened during the interview in Task 1. All 50 participants conversed with Interlocutor 1 (Int1) during Task 2, whereas half of the participants were interviewed by Interlocutor 2 (Int2) in Task 1 and the other half were interviewed by Interlocutor 3 (Int3). Int1 was found to have significantly lower click rates than Int2 and Int3 and their respective click rates were found to correlate with the click rates of the participants. Participants who clicked were shown to have a marked increase in click rate when interacting with Int2 and Int3 compared to when interacting with Int1. Although it was acknowledged that the Task 1 interactions may have provided more clicking opportunities than Task 2 and that there may have been an effect of gender, it was suggested that a plausible explanation for the findings could be that the participants' clicking behaviour was converging to that of their interlocutors.

In addition to the use of map-tasks, alternative sources of conversational speech data have included existing speech corpora such as the Wildcat Corpus (van Engen et al., 2010) and the Columbia Games Corpus (Beňuš, Gravano, & Hirschberg, 2007). For instance, Kim, Horton, & Bradlow (2011) used a selection of recorded conversations from the Wildcat Corpus to examine accommodation between pairs of native and non-native English speakers. The recorded conversations in this corpus were elicited using the interactive Diapix task (van Engen et al., 2010). In the Diapix task, pairs of interlocutors completed a 'spot the difference' task, where they were each presented with images which did not match, and their task was to find the ten ways in which their images differed. An updated version of the Diapix task has since been released known as the DiapixUK task (Baker & Hazan, 2011), which features an extended set of picture materials. The Wildcat Corpus did not contain pre-exposure baseline utterances or post-exposure utterances and therefore Kim et al. (2011) measured accommodation by comparing speech samples taken from early and late portions of the recorded conversations. A number of other accommodation studies have made use of the Diapix method (Alshangiti & Evans, 2011; Borrie & Delfino, 2017; Lewandowski & Jilka, 2019; Solanki, 2017; Solanki, Vinciarelli, Stuart-Smith, & Smith, 2015; Stamp, Schembri, Evans, & Cormier, 2015).

A range of accommodation studies have also analysed interactions from the Columbia Games Corpus (Heldner, Edlund, Hirschberg, & York, 2010; Levitan & Hirschberg, 2011; Nenkova et al., 2008). The Columbia Games Corpus contains recordings of twelve paired conversations elicited from native speakers of Standard American English. Participants played two computer games which involved one participant giving instructions to the other in order to identify objects and move them around the screen. After each game, the roles were reversed so that both participants experienced the role of the instruction giver and the instruction receiver. Participants were seated in the same soundproof booth but a curtain was used as a divider so that the participants could not see each other, or their respective computer screens. The aim of the game was to encourage collaboration and verbal discussion.

A major advantage of using collaborative speaking tasks to elicit speech for accommodation studies with a sociolinguistic focus, is that participants produce more natural, semi-spontaneous speech compared to read speech, whilst researchers are still able to maintain a high degree of control over the speech content that is produced. It could be argued that this is a relatively realistic version of what might happen in terms of accommodation behaviour in real life. However, the main drawback of using dyadic interactions compared to a non-interactive task is that the researcher cannot entirely control the input that the participants receive from their conversational partner, even if the same partner is present across all participants. Furthermore, Staum Casasanto et al. (2010) have noted that “no human confederate can entirely prevent his or her speech from being influenced by the naïve participant’s speech” (2010, p. 127). As previously discussed, one way that researchers have found to address this issue has been to analyse recordings of human-to-computer interactions where participants responded to a virtual interlocutor rather than a human partner. However, this method is considered less ecologically valid than human-to-human interactions as communication with a virtual interlocutor is likely to be less natural and representative of real-life scenarios than communication with a human being.

In almost all of the scenarios described so far in this section, the participants were not able to see the person they were communicating with; whether this be a model talker in a non-interactive, speech-shadowing task or a (human or virtual) partner in a conversational setting. The purpose of this was to prevent the use of facial expression, head nods, and other cues in order to ensure that participants were reliant on their speech to achieve successful communication. Although this seems like a logical choice for experimental purposes, in order to establish the extent to which speakers converge towards their interlocutors in natural, everyday conversations it seems more appropriate to allow interlocutors to have a face-to-face interaction. Although the effects of interlocutors being able to see each other on convergence rates has not been explicitly tested, studies have found that when faced with challenging speaking conditions, speakers adapted their speech to a stronger degree when they were unable to see each other than when they interacted face-to-face (Fitzpatrick, Kim, & Davis, 2015; Hazan et al., 2019). Based on this, it is possible that speakers may also converge towards their interlocutor to a lesser degree in situations where they are able to use other resources, such as gesture and gaze, to aid communication.

A final point that must be taken into account when considering the efficacy of using collaborative speaking tasks for accommodation research is that the nature of these tasks may influence accommodation behaviour. Levitan & Hirschberg (2011) noted that "the prevalence of entrainment in [the Columbia Games Corpus] may be attributable to the game domain, which lends itself to high levels of engagement on the part of the speakers." (2011, p. 3084). This sentiment was echoed by Schweitzer & Lewandowski (2014) when they stated that due to the fact that participants are required to communicate efficiently in order to successfully complete game tasks, "the question arises whether convergence in a game corpus may be a consequence of the game concept instead of a natural phenomenon in conversation." (2014, p. 391). It could be the case that rates of convergence are higher in speech elicited in the types of conversational settings that have previously been used in accommodation studies than we would find in natural, everyday casual conversations.

2.1.2.1.2.2 Casual conversations

Although the majority of studies investigating accommodation behaviour have involved laboratory tasks, some studies have analysed speech that has been elicited via casual conversations both in and out of the laboratory. In some cases, researchers have conducted interviews with their participants to elicit speech. For instance Natale (1975) examined convergence in vocal intensity using recordings of participants being interviewed by a researcher as well as having unstructured paired conversations with other participants. In this study it was found that the lowering or raising of the interviewer's vocal level produced a corresponding change in the vocal level of the participants. Gregory (1990) conducted face-to-face interviews with participants in order to analyse how f_0 was influenced by accommodation effects and found evidence of synchrony taking place in voices of interview partners. Coles (1993) also conducted face-to-face interviews with participants and organised participant group discussions to examine accommodation behaviour in /s/ variation. In this study, evidence of convergence was observed whereby so called 'semispeakers' of the Isleño dialect of Spanish were found to adapt their /s/ productions to become more similar to fluent speakers of the dialect. This can be classed as *upward convergence* as speakers considered to be of a lower status were found to adopt a feature of the prestigious dialect, typically spoken by speakers of a higher status.

A number of studies have elicited casual, conversational speech to examine accommodation without using an interview format. In some cases, participants were provided with particular topics to talk about and in other cases conversations were entirely unstructured. Bilous & Krauss (1988) gave their participants a brief written description of a problem and asked them to discuss it in pairs, come up with possible solutions and then arrive at a mutually acceptable agreement of how they would resolve the problem. Recordings of the conversations that took place were used to examine differences between men and women in accommodation behaviour across a wide range of conversational factors. Another source of data where participants were asked to talk naturally about a specific topic is the Switchboard Corpus (Godfrey, Holliman, & McDaniel, 1992) which contains recordings of approximately 2500 semi-spontaneous telephone conversations by 500 speakers from around the U.S. Of the 500 speakers who took part in this study, 50 "target" speakers participated in at least 25 telephone conversations, over a period of weeks, resulting in more than an hour of speech gathered for each of these participants. Each speaker was paired randomly by computer operator with various other speakers. For each conversation, participants were asked to discuss one of seventy possible topics with their interlocutor. Using a subset of telephone conversations from the Switchboard Corpus, accommodation studies have been conducted which have examined convergence in speech rate (Cohen Priva, Edelist, & Gleason, 2017; Fuscone, Favre, & Prévot, 2018) and entrainment in the use of high-frequency words (Nenkova et al., 2008), among other things.

Accommodation studies have also been conducted using entirely unstructured conversational data. For example, Purnell (2009) examined accommodation by black speakers in south-eastern Wisconsin toward white speakers, by recording pairs of participants talking about any topic of their choice, and found evidence of speakers accommodating in terms of vowel-quality position and diphthong elongation. Kurtić & Gorisch (2018) investigated accommodation in f_0 and turn competition in overlapping speech using data from a corpus of multi-party face-to-face conversations which took place between four British English speaking friends. These free flowing conversations were elicited without a task or a pre-assigned topic as part of a study conducted by Kurtić, Wells, Brown, Kempton, & Aker (2012). Some studies have also elicited their data by using a combination of casual conversations and collaborative speaking tasks.

In Manson, Bryant, Gervais, & Kline (2013), triads of same-sex strangers were asked to converse for ten minutes on any topic they wished. Convergence was assessed by comparing a number of features including speech rate, f_0, laughter counts, and language style from the start of the conversation to the end of the conversation. During these casual conversations, it was observed that participants tended to converge most in terms of speech rate (defined as average syllable duration) and language styles (calculated by measuring the proportion of words spoken that fall into nine particular function word categories). Following the casual conversations, participants were asked to take part in a short game. It was found that those participants who converged strongly in terms of speech rate during the casual conversation, were more likely to cooperate with their interlocutor during the game. However, participants who converged strongly in terms of language style matching score were no more likely to cooperate during the game.

There are also some examples of particularly innovative accommodation studies which have analysed data from real-life situations. For instance, one of the earliest examples is Coupland's (1984) study which examined conversations between a travel agent from Cardiff and 51 of her clients. In this study, the clients were grouped according to their occupational class and an analysis of four phonological variables revealed that the travel agent's use of these variables reflected the variation apparent in the clients' speech. Specifically, as percentages of the less standard variants increased in the clients' speech, they also increased in the travel agent's speech. In a more recent study, Lee et al. (2010) investigated accommodation in married couples' problem-solving interactions which took place during couples therapy. The couples were asked to discuss problems in their relationship and then they were recorded having casual conversations, without a therapist or member of research staff present. Lee et al. (2010) found higher levels of coordination in terms of f_0 in those couples whose interactions were considered to be positive than those who had negative interactions. Another study which considered how accommodation behaviour related to relationship stability was conducted by Ireland et al. (2011). In this study, participants went on four minute speed dates with up to twelve opposite-sex individuals and the casual conversations that took place were recorded and analysed. Instant messages of those participants who chose to remain in contact after the speed dates were also analysed so that both relationship initiation and stability could be taken into account. It was found that outcomes for romantic relationships could be predicted based on how much participants' language styles matched in terms of their use of function words.

Language style matching was also analysed in an accommodation study which made use of transcripts of dialogue from nine actual hostage crises (Taylor & Thomas, 2008). Negotiation incidents were classified as being successful or unsuccessful based on whether the negotiation generated a peaceful outcome. Using these classifications, the relationship between language style matching and negotiation outcomes was analysed. It was found that correlations between the police negotiator and the hostage taker, in terms of language style at the conversational level, were generally higher in successful negotiations than in unsuccessful negotiations. However, at the turn-by-turn level a much more complex pattern was observed. Studies such as this demonstrate that it is possible to analyse speech accommodation using live, interactional data in order to establish how speakers behave in real-life scenarios.

In terms of ecological validity, the findings of accommodation studies which make use of this type of casual speech data can be said to be of great value. This type of data can provide an increased opportunity for researchers to generalise their results to real-world naturally occurring interaction. However, the main disadvantage of using casual conversations to analyse accommodation is that its spontaneous nature can make it more challenging to process and analyse compared to non-interactive speech data collected during speech-shadowing tasks, for example. Furthermore, a high degree of ecological validity often comes at the expense of experimental control; the researcher cannot control the input of other interlocutors or ensure that the speaker of interest produces an adequate amount of the specific types of sounds, words or structures required for the analysis. Figure 2.1 illustrates the hierarchy of the different types of speech data discussed in this section, in terms of ecological validity and experimental control. It is considered that the higher the experimental control, the lower the ecological validity and vice versa.

Experimental Control	**Data**	**Ecological Validity**
High	Non-interactive task	Low
	Human-to-computer interactive task	
	Human-to-human interactive task	
	Human-to-human casual conversation (laboratory)	
Low	Human-to-human casual conversation (real-world)	High

Figure 2.1. Hierarchy of speech data in terms of experimental control and ecological validity.

Pardo et al. (2018b) have suggested that phonetic convergence is not directly compatible across non-interactive and conversational settings and that “findings from non-interactive tasks, like speech shadowing, might transfer to conversational settings when considering average levels of convergence, but the lack of consistency for individual talkers across settings indicates that other talker-modulated factors might not transfer across settings.” (2018b, p. 8). If it is the case that patterns and mechanisms revealed using non-interactive tasks do not easily transfer to conversational interaction, then it is crucial that further studies into speech accommodation are conducted using more naturalistic, conversational data. This is particularly important in the investigation of accommodation behaviour presented in the present thesis, as one of its primary purposes is to establish how much speakers might adapt their speech across forensically-relevant speaking situations and to assess the implications that this might have on FSC analyses. Consequently, in order to ensure that the results of the study are applicable in real-world situations, greater importance is placed on ensuring that the data is forensically-relevant (i.e. ecological valid) rather than the focus being on experimental control.

The data analysed in the present thesis is from the West Yorkshire Regional English Database (WYRED; Gold, Ross, & Earnshaw, 2018) and all recordings are of studio quality, elicited in a laboratory setting. All of the speech is semi-spontaneous, insofar as it is not read speech but cannot be considered to be truly spontaneous due to the laboratory setting in which it is recorded, and it has been elicited in what are considered to be forensically-relevant speaking situations. A combination of human-to-human, face-to-face interactive tasks and casual conversations are analysed across three different scenarios.

The first scenario involves a simulated police interview in which participants are required to provide certain information whilst denying anything incriminating. This task is an extension of the HCRC map task technique (Anderson et al., 1991) which was modified for the DyViS project (Nolan et al., 2009) to include the roles of the police interviewer and the suspect, and then further adapted for the purpose of the WYRED project (Gold et al., 2018). This data is therefore most comparable to the DyViS data used in Gold et al.'s (2013a) investigation of accommodation in clicking behaviour but it is also considered to be similar in nature to other accommodation studies involving conversations elicited as part of an interactive task. Using the scale shown in Figure 2.1, this data would fall into the "human-to-human interactive task" category and therefore it is considered to have a relatively good balance between ecological validity and experimental control.

The second scenario considered in this investigation involves pairs of participants having a casual conversation with one another. Participants are given prompt cards providing them with suggested topics of conversation but are advised that they can talk about anything that they wish and therefore these conversations are considered to be examples of unstructured, casual conversations similar to those used in studies such as Purnell (2009) and Kurtić & Gorisch (2018). The final scenario involves participants leaving an answer message, guided by a short list of points that they need to convey in the message. These two tasks are believed to be more realistic representations of what might happen in real life compared to the first task; however, it is considered that the simulated police interview does elicit semi-spontaneous speech in a relatively stressful situation, despite the laboratory based experimental set-up. Using the scale shown in Figure 2.1, this data would fall into the “human-to-human casual conversation (laboratory)” category and therefore it is considered to have a high degree of ecological validity but less experimental control than some of the previous speech accommodation studies. Full details of the data used in this investigation can be found in Chapter 3.

2.1.2.2 Analytical approaches

The previous sub-section has provided an account of the types of data that have been used to examine speech accommodation in existing research. This sub-section presents an overview of the two main ways in which accommodation is typically measured: perceptually and via acoustic-phonetic analysis. This is presented alongside a summary of the advantages and disadvantages of these two approaches, which helps to justify the approach that has been taken in the present thesis. Firstly, some of the commonalities in what both types of accommodation study set out to achieve are described.

A fundamental aim of an accommodation study is to establish if and how a person’s speech changes as a result of exposure to a model talker or interlocutor. In both perceptual and acoustic-phonetic accommodation studies the analyst compares the participant’s speech at different points in time. In non-interactive speech-shadowing studies this typically involves comparing baseline productions of a set of target items with shadowed utterances of those same items and these productions are subsequently considered in relation to those of the model talker. This is also the case for some studies involving interactive conversation, in cases where participants are required to produce pre-exposure baseline utterances, shadowed utterances and sometimes post-exposure utterances (Bailly & Martin, 2014; Cao, 2018; Mukherjee et al., 2017; Pardo et al., 2013a; Staum Casasanto et al., 2010).

One important difference between comparing baseline productions with shadowed speech from a speech-shadowing task, versus shadowed speech from an interactive conversation is that the latter typically involves comparing different speaking styles. In most cases, baseline productions are taken from read speech elicited via a formal reading task involving a word list, whereas the shadowed speech elicited during an interactive speaking task or casual conversation tends to involve (semi-)spontaneous speech. Labov theorised that language can be organised along a single dimension, measured by the amount of attention paid to speech; whereby casual speech is placed at one end of the continuum and reading tasks are placed at the other (1972, p. 208). Labov asserted that in instances where greater attention is paid to speech more standard forms will tend to be used, whereas when less attention is paid to speech more vernacular styles will be used. It is therefore necessary to be aware that some differences may arise between the baseline and shadowed productions as a consequence of mismatch in speaking style, as well as the interlocutor's influence. In the present thesis, all of the speech analysed is semi-spontaneous rather than being elicited via reading tasks. However, levels of formality do vary across the different speaking tasks, with more standard forms being expected during the simulated police interview than in the other two tasks. Therefore, the results of the short-term accommodation case studies presented in this thesis are discussed with reference to differences in speaking style and the interlocutor, although it is not possible to disentangle these factors from one another.

Other studies have investigated accommodation without eliciting baseline productions prior to exposure to a model talker/interlocutor. In these instances, levels of accommodation are sometimes assessed by comparing utterances from different time points within a single interaction. For instance, Levitan & Hirschberg (2011) measured convergence at the conversation level in a range of acoustic features by using paired t-tests to compare productions from the first half of the paired conversation with productions from the second half of the conversation. In cases where the distances between interlocutors were smaller in the second half than the first, convergence was inferred. Similarly, Kim et al. (2011) assessed accommodation by comparing productions of tokens from the first third of a conversation with tokens from the final third, using a perceptual method of analysis. Lewandowski & Jilka (2019) also measured convergence in amplitude between pairs of speakers by dividing their conversations into thirds and comparing amplitude envelopes from the first third of the conversation with amplitude envelopes from the last third.

By measuring accommodation over the course of a single interaction, it is possible to consider how quickly any accommodative behaviour takes place and this also avoids any complications that might arise by comparing different speaking styles across baseline and shadowed conditions. However, it must be acknowledged that in some cases shifts in speech have been found to pattern with changes in conversational topic over the course of a single interaction. For example, Love & Walker (2013) found evidence of a modest topic effect whereby participants had higher rates of rhoticity (a salient feature of American English) when talking about American football teams compared to when talking about English Premier League football clubs. A range of other studies have also reported robust effects of topic on linguistic variation when comparing work and non-work related topics (Devlin, 2014; Douglas-Cowie, 1978; Leach, 2018).

The first methodological approach to be described below is the use of perceptual similarity ratings. The second approach involves the use of acoustic-phonetic analysis. A greater emphasis has been placed on describing the acoustic-phonetic approach as this is the approach that has been applied in the research presented in this thesis. However, a brief description of the perceptual approach is considered to be useful for the reader as it illustrates how evidence has been gathered to demonstrate that short-term accommodation can be perceived by listeners as well as being measurable via auditory and acoustic analysis.

2.1.2.2.1 Perceptual

The perceptual approach to measuring speech accommodation involves collecting holistic evaluations of speech samples from listeners in order to determine whether speakers are judged to sound more similar to each other after interacting with one another than prior to exposure. Goldinger (1998) was the first scholar to adapt a classic psychophysical AXB perceptual similarity paradigm to examine phonetic convergence via a speech-shadowing task. This methodology involved a set of naïve listeners being asked to judge whether participant productions made during a baseline task (A) or participant productions made during a shadowing/post-exposure task (B) sounded like a better imitation of productions made by the model talker (X). On each trial, listeners were presented with three versions of the same item with the A and B stimuli surrounding the X stimulus (with baseline tokens counterbalanced across the first and third positions across groups). If the production made during the shadowing/post-exposure task were judged to be a better imitation of the model talker than the productions made during the baseline task, then this could be taken to indicate that convergence had taken place.

A large number of other accommodation studies have since applied a similar approach to Goldinger (1998) by using AXB perceptual similarity measures to assess convergence levels. Many such studies have involved non-interactive speech-shadowing tasks (Babel et al., 2014; Namy et al., 2002; Pardo et al., 2018b; Shockley et al., 2004) while others have analysed interactive conversation using a perceptual approach (Alshangiti & Evans, 2011; Kim et al., 2011; Pardo, 2006; Pardo et al., 2018b). Most studies have slightly adapted the instructions to listeners in order to ask them to judge similarity in pronunciation rather than judging imitation. However, Pardo et al. (2010) found no differences in findings according to whether listeners judged imitation or similarity in pronunciation.

Goldinger (1998) suggested that this holistic approach to examining phonetic convergence better reflected perceptual similarity between tokens and provided more valid measures than imitation scores derived from acoustic measures of specific phonetic parameters, as it was said that "imitation is in the ear of the beholder" (1998, p. 257). Using a perceptual method of analysis, listeners are able to make an assessment that takes into account multiple acoustic parameters in the speech signal in order to form a holistic judgement of similarity. Some scholars have argued that these ratings can be more useful for drawing broad conclusions regarding speech accommodation than acoustic assessments of individual phonetic parameters (Pardo et al., 2017). On the other hand, the main limitation of the AXB procedure is that it does not provide an opportunity to explore in what respect a speaker has accommodated. The listeners could be responding to any number of variables and therefore if it is the aim of the researcher to establish specifically the ways in which a person adapts their speech as a result of accommodation effects, an acoustic-phonetic approach may be more desirable.

In the early work by social psychologists, prior to the implementation of the AXB paradigm, accommodation would generally be measured impressionistically by asking by naïve listeners to rate aspects such as how 'broad' the speakers' accents sounded. It was around this time that Trudgill (1986, p. 4) outlined a number of motivations for incorporating detailed linguistic analysis into the study of speech accommodation. Specifically it was proposed that detailed linguistic analyses would permit, amongst other things:

1. An exact, rather than impressionistic, quantification of degree of linguistic accommodation;
2. An examination of which linguistic features are and are not changed during accommodation, together with explanations for this;
3. A study of whether accommodation is a uniform process, or whether linguistically different types of accommodation take place in the case of different speakers, different situations, or different relationships;
4. A study of the limits of accommodation: what are the linguistic (as opposed to social and psychological) constraints on accommodation, and is it possible to accommodate totally to a new variety?

   (Trudgill, 1986, p. 4)

For researchers who aim to address any of the above points, an acoustic-phonetic approach to analysing speech accommodation may be preferable over a perceptual approach, or alternatively a combination of the two approaches may be desirable.

2.1.2.2.2 Acoustic-phonetic

The acoustic-phonetic approach to assessing speech accommodation involves analysing individual acoustic parameters within the speech of different interlocutors, in order to quantify the degree and direction of any changes that may occur either during or after their interaction. This approach has been used to analyse data from interactive conversations as well as speech-shadowing tasks, to a slightly lesser degree. The precise technique for measuring speech accommodation can vary across different acoustic-phonetic studies depending on the type of data that is being analysed and how accommodation is defined.

For the purposes of measuring speech accommodation using an acoustic-phonetic approach, convergence has been defined in numerous ways. For instance, some studies have contrasted entrainment at the conversation level with entrainment at the turn level (Edlund, Heldner, & Hirschberg, 2009; Levitan & Hirschberg, 2011). In these studies, the term *entrainment* is used to refer to the general phenomenon whereby people engaging in conversation become more similar to their interlocutor as the conversation proceeds. They then break entrainment down into the initial aspect of *proximity* (i.e. how similar speakers are) which is then followed by *convergence*, which they define as “an increase in proximity over time” (Levitan & Hirschberg, 2011, p. 3081). They also identify another property at the turn level, known as *synchrony*, which refers to “a turn-by-turn relative coordination between partners” (Levitan & Hirschberg, 2011, p. 3081). Slightly different definitions and distinctions have been applied in other studies. For instance, Kurtić & Gorisch (2018) assess f_0 accommodation at the conversation level and at the turn level using two separate approaches referred to as *initialisation* and *normalisation*. In this study, initialisation is defined as “f_0 accommodation to co-participants’ or speakers’ own values in turns containing and preceding overlap” and normalisation refers to “speakers’ f_0 accommodation relative to their individual norms generalised over the entire conversation” (Kurtić & Gorisch, 2018, p. 377). As accommodation is defined and measured in different ways across different studies, care must be taken when comparing findings from multiple studies.

As speech accommodation involves speakers adapting their speech according to whom they are speaking, many acoustic-phonetic assessments of this phenomenon involve an examination of the parameter under investigation over the whole duration of the task or interaction, so that any changes over time can be tracked. In cases where clear evidence of convergence have been observed, it is often the case that the degree of convergence at the start of the interaction is lower than at the end (Levitan & Hirschberg, 2011; Pardo, 2006). In order to measure and assess accommodation using an acoustic-phonetic approach, it is also necessary to establish how the parameter of interest is produced by the model talker/interlocutor. Various techniques for analysing the speech of the interlocutor have been applied across different studies.

In speech-shadowing studies, acoustic measures are taken from the pre-recorded stimuli produced by the model talker and direct comparisons can be made between participant and model talker productions of each item. The presentation order and the frequency of each item being presented during the speech-shadowing task are usually carefully controlled and recorded so that the effects of these factors can be taken into account. In accommodation studies involving interactive conversations, acoustic measurements are sometimes taken from the interlocutor's speech during the interaction and in other cases measures are collected from an established baseline for the interlocutor. In studies where the interlocutor's productions during the paired conversation are analysed, either direct comparisons are made between participant and interlocutor productions of each item, or an average acoustic value is calculated for the interlocutor across multiple items and then a set of distances are measured between the participant's tokens and the interlocutor's average value (Cao, 2018).

In any experimental design involving human-to-human interactive conversations, it is possible that both the participant and the interlocutor may display accommodation behaviour. For this reason, some studies have measured the participants' accommodation behaviour relative to their interlocutor's baseline productions. For example, Cohen Priva et al. (2017) measured the baseline speech rate of each participant's interlocutor by calculating the interlocutor's average speech rate when they were not interacting with the participant but were instead conversing with a number of other speakers. Using this measure, the interlocutors' baseline speech rate was found to be a significant predictor of the participants' speech rate, with participants' speech rate changing in the same direction as their interlocutor when speaking to an interlocutor who spoke slowly or quickly. While this technique is a useful way of establishing a set of typical acoustic values for a particular speaker, this is only possible when recordings of each speaker talking to multiple other speakers exist. The Switchboard Corpus is one such example of a database containing this type of data (Godfrey et al., 1992).

Acoustic-phonetic accommodation studies vary greatly not only in terms of how accommodation is measured but also in terms of how many participants and model talkers/interlocutors are involved. Additionally, a wide range of different speech parameters have been assessed in terms of how they are influenced by speech accommodation. Table 2.1 provides a summary of the phonetic variables that have been examined in a selection of previous acoustic-phonetic accommodation studies published in the last decade. A large proportion of these studies involve human-to-human interaction as acoustic-phonetic studies tend to involve this type of data to a larger degree than perceptual studies do. Studies have been organised firstly according to the phonetic feature under investigation and secondly by chronological date order. Studies investigating more than one parameter are listed at the end of the table. It should be noted that some studies used a combined approach where the acoustic-phonetic analysis was supplemented by perceptual similarity ratings (e.g. Babel & Bulatov, 2012; Cao, 2015; Pardo et al., 2010, 2017; Walker & Campbell-Kibler, 2015 *inter alia*).

The summary presented in Table 2.1 is by no means intended to be an exhaustive list of all of the studies which have examined accommodation using an acoustic-phonetic approach; rather it is intended to highlight the wide range of parameters that have been shown to be influenced by speech accommodation. Details of the number of participants included in the studies are provided to illustrate the range in scale across different investigations. Information about the type of data used in each study as well as details of with whom the participants interacted/shadowed are also presented. Details of where the participants and their interlocutors/model talkers are from, as well as their respective genders are included as a number of acoustic-phonetic accommodation studies have focused on the influence of where the interlocutors are from and also the effects of gender (Namy et al., 2002; Pardo, 2006; Pardo et al., 2010). Factors such as age and social class are not included in this summary table as this information was not always as readily available.

Accommodation can occur over various language levels including lexical choices, syntactic structure, prosodic features and phonetic parameters. The summary presented in Table 2.1 focuses particularly on segmental and suprasegmental phonetic features that have been analysed in previous accommodation studies. However, it should be noted that accommodation studies into lexical and syntactic alignment follow the same general principles as acoustic-phonetic studies, insofar as researchers carry out linguistic analyses to determine how the parameter of interest is influenced by accommodation, rather than employing naïve listeners to make these judgements. Some studies have investigated lexical alignment (Branigan et al., 2011; Cowan et al., 2019; Cowan & Branigan, 2015; Garrod & Doherty, 1994), and others have examined syntactic alignment (Bock, 1986; Cowan et al., 2015; Weatherholtz et al., 2014). Studies selected for inclusion in this summary are mainly ones where the primary focus was to investigate speech accommodation.

Table 2.1. Summary of acoustic-phonetic accommodation literature organised by acoustic parameter.

Parameter	Reference	Data type	Participant details	Model talker/interlocutor details
Amplitude/Intensity	Lewandowski & Jilka, 2019	Human-to-human interactive task	20 native German speakers (10 female, 10 male)	American English male & SSBE female interlocutors
Clicks	Gold et al., 2013a	Human-to-human interactive task	50 SSBE male speakers	SSBE male & Dutch female/Australian female interlocutors
Fundamental frequency (f_0)	Heldner et al., 2010	Human-to-human interactive task	13 American English speakers (6 female, 7 male)	Participant pairings
	Babel & Bulatov, 2012	Non-interactive task	19 American English speakers (12 female, 7 male)	American English male model talker
	Gijssels et al., 2016	Human-to-computer interactive task	72 native Dutch speakers (48 female, 24 male)	2 Dutch model talkers (virtual; male - "VIRTUO", female - "VIRTUA)
	Mukherjee et al., 2017	Human-to-human interactive task	16 native Italian speakers (8 female, 8 male)	Same-sex participant pairings
	Kurtić & Gorisch, 2018	Human-to-human casual conversation (laboratory)	4 British English speakers (3 female, 1 male)	Participant grouping
	Ibrahim et al., 2019	Human-to-computer interactive task	60 Swedish speakers (30 female, 30 male)	Swedish male model talker (virtual - "Furhat")
Mel-frequency cepstral coefficients (MFCC)	Bailly & Martin, 2014	Human-to-human interactive task	35 pairs of French speakers	Mixed and same-sex participant pairings
Speech rate/Articulation rate	Staum Casasanto et al., 2010	Human-to-computer interactive task	62 native Dutch speakers (32 female, 30 male)	Dutch male model talker (virtual - "VIRTUO")
	Cohen Priva et al., 2017	Human-to-human casual conversation (laboratory)	481 American English speakers	Mixed and same-sex participant pairings
	Fuscone et al., 2018	Human-to-human casual conversation (laboratory)	479 American English speakers	Mixed and same-sex participant pairings

Parameter	Reference	Data type	Participant details	Model talker/interlocutor details
Voice onset time (VOT)	Nielsen, 2011	Non-interactive task	25 American English speakers (12 female, 13 male)	Phonetically trained American English male model talker
	Yu et al., 2013	Non-interactive task	84 American English speakers (45 female, 39 male)	Native English male model talker
	Solanki, 2017	Human-to-human interactive task	12 Scottish English female speakers	Same-sex participant pairings
Voice quality	Borrie & Delfino, 2017	Human-to-human interactive task	20 American English female speakers	2 female American English interlocutors
Vowel spectra	Babel, 2010	Non-interactive task	42 New Zealand English speakers (32 female, 8 male)	Australian English male model talker
	Babel, 2012	Non-interactive task	107 American English speakers	2 American English male model talkers (1 black, 1 white)
	Pardo et al., 2012	Non-interactive task	10 native English male speakers	College roommate pairings [long-term accommodation]
	Cao, 2015	Human-to-human interactive task	16 Mandarin L2 English female speakers	Australian English female model talker
	Walker & Campbell-Kibler, 2015	Non-interactive task	36 English female speakers (20 from New Zealand, 16 from Columbus Ohio)	4 female model talkers (New Zealand, Australia, U.S. Inland North and U.S. Midland)
Amplitude/Intensity, f_0, Speech rate/Articulation rate & Voice quality	Levitan & Hirschberg, 2011	Human-to-human interactive task	13 American English speakers (6 female, 7 male)	Participant pairings
f_0 & Intonation	Lee et al., 2010	Human-to-human casual conversation (real-world)	120 American English speakers (60 married couples)	Married partner pairings

Parameter	Reference	Data type	Participant details	Model talker/interlocutor details
f_0, Intonation & Allophonic contrasts	Gessinger et al., 2019	Human-to-computer interactive task	12 native German speakers (9 female, 3 male)	German female model talker (virtual - "Mirabella")
f_0, Vocalic duration & Vowel spectra	Pardo et al., 2017	Non-interactive task	92 native English speakers (47 female, 45 male)	12 native English model talkers (6 female, 6 male); mixed & same-sex pairings
	Pardo, Jordan, Mallari, Scanlon, & Lewandowski, 2013	Non-interactive task	20 native English speakers (10 female, 10 male)	20 native English model talkers (10 female, 10 male); same-sex pairings
Filled pauses & Speech rate/Articulation rate	Pardo et al., 2013a	Human-to-human interactive task	16 native English speakers (8 female, 8 male)	Same-sex participant pairings
Speech rate/Articulation rate & turn-taking	Finlayson et al., 2012	Human-to-human interactive task	64 speakers (32 female, 32 male; 61 Scottish, 2 English, 1 American)	Participant pairings
Speech rate/Articulation rate & Vowel spectra	Pardo et al., 2010	Human-to-human interactive task	24 American English speakers (12 female, 12 male)	Same-sex participant pairings
Speech rate/Articulation rate, f_0, Language Style & Laughter analysis	Manson et al., 2013	Human-to-human interactive task & casual conversation (laboratory)	105 native English speakers (60 female, 45 male)	Same-sex participant triads
Fricatives /z/ and /θ/, Rhoticity & Vowel spectra	Cao, 2018	Human-to-human interactive task	19 Hong Kong English speakers (12 female, 7 male)	4 female model talkers (2 American English, 2 RP English)

As previously discussed, one of the main advantages of the acoustic-phonetic approach to analysing speech accommodation is that this method enables the researcher to examine how specific parameters of interest are influenced by accommodation. However, a disadvantage of this approach is that it is generally much more time consuming to analyse features individually compared to assessing accommodation using perceptual similarity judgements. For this reason, many acoustic-phonetic studies tend to focus on only a single parameter or a small selection of individual variables. In relation to this, some scholars have argued that the question of whether one speaker has converged towards another speaker cannot be adequately addressed with reference only to a selection of individual acoustic parameters.

Pardo et al. (2017) suggested that "assessments of a single acoustic attribute are limited with respect to broader interpretations of the phenomenon." (2017, p. 642). This echoed the sentiments of Goldinger (1998) where it was stated that an acoustic-phonetic approach to assessing convergence may miss the "perceptual Gestalt" (1998, p. 257). A key principle of Gestalt theory is that 'the whole is greater than the sum of its parts' and therefore this "perceptual Gestalt" can be considered as a 'holistic', 'global' or 'overall' perception of convergence. In studies which have found evidence of convergence with respect to an individual feature using acoustic methods, it is not necessarily known whether this convergence would also be perceived by listeners. Some studies have attempted to examine this by using a combined acoustic-phonetic and perceptual approach to assess levels of convergence. For instance, Pardo et al. (2013b) justified their use of a combined approach by stating that "perceptual measures provide a global estimate of convergence, while acoustic measures contribute to an understanding of the attributes that talkers employ when converging." (2013b, p. 185).

It is also important to recognise that although listeners may perceive speakers to have either converged or diverged, it is often the case that accommodation behaviour can vary across different speech parameters which can make it challenging to present an overall view of whether convergence has taken place, using an acoustic-phonetic only analysis. For instance, studies which have analysed multiple acoustic parameters have sometimes reported inconsistencies across measures of phonetic convergence in both conversational interaction and in speech-shadowing tasks. Pardo et al. (2018b) noted that in the previous studies by

Pardo et al. (2013b) and Pardo et al. (2017), measures of duration, f_0, F1, and F2 each exhibited "a distinct, talker and item-dependent pattern of variation and convergence." (2018, p. 3). In both of these speech-shadowing studies, examinations of each measure alone revealed different patterns of results across different measures. For example, it was found that a talker might converge only in duration, or in duration of some items but in vowel formants of other items (Pardo et al., 2018b, p. 3). Bilous & Krauss (1988) also found a similar trend, whereby speakers were observed converging in some parameters whilst simultaneously diverging in others. A likely reason for the inconsistencies in accommodation behaviour identified in studies such as these is that certain parameters may be more susceptible to accommodation than others.

It was suggested by Trudgill (1981) that attention and social salience are key factors in explaining phonetic convergence. Building on this idea, Babel (2009) argued that speakers are likely to display stronger accommodation effects in phonetic parameters for which they have a wider variety of stored representations (pp. 56–57). Walker & Campbell-Kibler (2015) observed patterns in line with this whereby convergence appeared to be promoted by larger phonetic differences between the model talker and the shadower, as well as greater existing variability in a vowel class. Additionally, Cao (2018) found evidence of speakers tending to converge most strongly on linguistic features where the greatest initial differences were present between the participants' realisations and those of the model talkers. Of the five phonetic features that were explored in this study, rhoticity was found to be the feature on which speakers converged to the strongest degree overall and this feature was deemed to be most salient to them.

Smith & Holmes-Elliott (2015 & 2017) also found evidence of accommodative behaviour being influenced by how similar the participants' speech was to their respective interlocutors. In Smith & Holmes-Elliott (2015), speakers from Buckie (on the North East coast of Scotland) were recorded having separate conversations with a community insider (i.e. someone from Buckie) and a community outsider (someone from outside of the local area). In cases where participants conversed with community outsiders, where the greatest differences between interlocutors were present, participants were found to shift towards standard variants, but only with respect to variables that were socially salient and/or stigmatised in the community.

One of the primary purposes of the research presented in the present thesis is to examine how two socially salient phonetic variables change across different forensically-relevant scenarios, involving different speaking styles and interlocutors. Consequently, an acoustic-phonetic approach is considered to be more suitable than a perceptual approach, for the reasons outlined above. Although it is acknowledged that a supplementary perceptual analysis could complement the acoustic-phonetic analysis and provide a more holistic view of any accommodative behaviour, this is beyond the scope of the present investigation.

2.2. Practical implications of accommodation

This section describes some of the practical implications of accommodation for researchers in the fields of sociolinguistics and forensic speech science and also briefly outlines how accommodation research can contribute to the development of spoken dialogue systems.

2.2.1. Sociolinguistics

Although there is debate as to whether speech accommodation occurs unconsciously as a result of automatic, cognitive processes, or whether it is socially motivated and consciously controlled, it is clear that convergence in particular can function as a resource to aid communication and build rapport. It would seem that convergence can help to achieve efficient communication, as Nenkova et al. (2008) found that conversations which had high levels of convergence on high-frequency words were more likely to rate highly in terms of naturalness, coordinated turn-taking behaviour and overall task success. Other studies have shown that convergence can help to reduce misunderstanding (Pickering & Garrod, 2013) and be used as a means to display a positive attitude towards an interlocutor (Lee et al., 2010). Furthermore, Giles et al. (1987) note that convergence to another's dialect can cause people to attribute a range of positive traits to the person who has converged, such as friendliness and warmth (1987, p. 15). Zellers & Schweitzer (2017) have also found evidence of higher rating of conversational friendliness in interactions where participants have converged towards one another. It is for the above reasons that accommodation is considered to be a useful conversational strategy that sociolinguistic researchers should be aware of when analysing social interaction.

It is possible that when conducting sociolinguistic interviews, researchers may find that by converging towards their participants in some respects, the participants may be more responsive to their questions and communication may be more effective. Related to this, researchers must also acknowledge that their own language use may influence that of their participants and therefore attention should be paid to this potential confounding factor during the data collection process. Equally, researchers should be aware that participants recorded as part of the same conversation or task may influence each other's linguistic behaviour, and therefore may not be entirely independent of one another. Whether the focus of a sociolinguistic study is to examine speech accommodation or not, it may still be useful to consider how accommodative behaviour may shape the interaction.

Another implication of speech accommodation is the potential for language to evolve over time as a long-term consequence of this phenomenon. It has been suggested by a number of scholars that convergence is a possible mechanism for propagation of sound change (Auer & Hinskens, 2005; Babel et al., 2014; Delvaux & Soquet, 2007; Tamminga, 2016; Trudgill, 1986). Auer & Hinskens (2005) describe the change-by-accommodation model which suggests that "the driving force of language change is interpersonal accommodation (convergence)" (2005, p. 356). The findings of their study suggest that "the driving force behind change in the individual, and also in the community, is not imitation of the language of one's interlocutor but, rather, an attempt to assimilate one's language to the possibly stereotyped characteristics of a group one wants to be part of, or resemble." (2005, p. 356). Delvaux & Soquet (2007) also discuss the potential effects of multiple imitative speech interactions on sound change and suggest that when speakers converge "not only are individual phonetic realisations modified, but mental (phonetic) representations are also" (2007, pp. 3–4). They conclude that the results of their psycholinguistics study support the theory that speech imitation is one of the bases of sound change.

2.2.2. Forensic speech science

It is also important for scientists in other areas of linguistic research, such as forensic speech scientists involved in undertaking FSC casework and language analysis for the determination of origin (LADO), to acknowledge the various implications that speech accommodation could have on the speech that practitioners in these areas encounter during their analyses. To the author's knowledge, no accommodation studies to date have explicitly discussed the implications of speech accommodation for FSC casework or LADO assessments. However, there have been some studies that have considered the relationship between accommodation and attribution of guilt in legal settings more generally.

Dixon, Tredoux, Durrheim, & Foster (1994) examined the attribution of guilt as a function of speech accommodation and crime type. Based on listener ratings, it was found that the Cape Afrikaans-speaking criminal suspect was judged to be significantly less guilty when converging towards the English-speaking interrogator as opposed to diverging into Cape Afrikaans. While the findings of this study reflected a preference for convergent behaviour, it must also be acknowledged that underlying biases related to the Cape Afrikaans accent may have been at play. Indeed, Dixon et al. (1994) recognised the possibility that "the affective bases of speech accommodation are less relevant in legal contexts than are the prestige and well-formedness of the vernacular selected by suspects" (1994, p. 472). Nevertheless, this study provided empirical evidence to demonstrate that speech accommodation, and more specifically how the consequences of this phenomenon are perceived, can have very real and important implications in forensic settings.

2.2.2.1 Forensic speaker comparison (FSC)

In the context of forensic speech science research, experts must consider what impact any potential effects of speech accommodation might have on the outcome of a FSC case. At present, although the expert can take into account any mismatches between evidential samples in terms of channel, speaking mode or stylistic variability when interpreting their findings, it is difficult for them to consider any potential influence of the interlocutor in detail. This is partly because one of the initial preparatory stages in FSC casework is to produce edit samples of the speaker of interest in the reference and questioned sample(s) by removing any other voices and portions of excessive background noise from the original recordings (Foulkes & French, 2012, p. 7). These working samples are ideally prepared by someone other than the expert carrying out the comparison which means that experts rarely consider the speech of any interlocutors present in the evidential samples. In this thesis, accommodation behaviour is examined to explore how much certain parameters can vary within an individual in forensically-relevant contexts and, based on these findings, it is also considered whether it would be useful to routinely examine the speech of interlocutors as part of a FSC.

It could be argued that in some extreme cases of short-term accommodation, speech samples can become “cross-contaminated” by the interlocutor, as features of the interlocutor’s speech can be reflected in the criminal or suspect’s speech, as a result of convergence. Or conversely, a speaker might deviate far from their typical way of speaking in order to diverge away from their interlocutor. Either of these types of accommodation behaviour have the potential to result in differences arising between the evidential samples even in cases where they are produced by the same speaker. For instance, it is possible to imagine a scenario where someone with what might be described as a strong regional accent might display many more regional features when interacting casually with a friend or accomplice from the same area, with a similar accent, compared to when interacting formally with a police officer with a less regionally marked accent. Consequently, this could lead to a number of segmental differences between samples. Although this would not necessarily have a strong impact on the overall conclusion if there were a range of other features that could be analysed which were not affected by accommodation, in cases involving poor quality material where only a limited range of parameters could be analysed, this could have a more substantial impact on a FSC case. It is also possible that accommodation could occur in a range of different parameters (e.g. voice quality, f_0, certain vowels and consonants) resulting in very strong differences between evidential samples of the same speaker.

Often in FSC cases where it is thought that a speaker is disguising their voice or imitating someone else, the expert will be unable to conduct a reliable analysis, as it will not be possible to determine the underlying features of the speaker. In much the same way as it is necessary for an expert to exercise caution if there is evidence to indicate that a speaker has deliberately attempted to disguise their speech, it is possible that in cases where a speaker appears to be accommodating to a very strong degree this may also render a comparison unreliable, leading to the case being screened out altogether. Alternatively, it may be the case that an analysis can be attempted but conclusions may be more limited in instances where high levels of variability are present that may be explained by accommodation behaviour.

Evidence has been presented throughout this chapter to demonstrate that speakers can accommodate in numerous ways across a vast range of phonetic and non-phonetic parameters. However, the majority of previous accommodation studies have involved the analysis of speech which has been collected in situations which are not considered to be directly relevant for FSC casework. It is therefore important that we build on this knowledge to understand the extent to which parameters can vary within an individual in forensically-relevant contexts involving different interlocutors and speaking styles, as it may be the case that within-speaker variability is even stronger when both of these factors differ. The research presented in this thesis aims specifically to address this issue by examining how speakers' realisations of two socially salient parameters vary across a range of different speaking tasks that have been designed to be as forensically-relevant as possible (in a laboratory setting). The findings of this research will be discussed within the context of FSC casework.

2.2.2.2 Language analysis for the determination of origin (LADO)

Another practical setting in which speech accommodation could have important implications is that of immigration cases where the claimed nationality of an asylum seeker is in dispute. In many cases where a person's nationality cannot be reliably evidenced, immigration services have used language analysis as a tool to investigate the claimed origin of the asylum seeker (Cambier-Langeveld, 2014). In these cases, an analyst will typically conduct an interview with the claimant in order to obtain speech evidence that can be analysed to provide additional evidence in the case. A prominent debate within the field of LADO research has surrounded the question of whether non-linguist native speakers can conduct these analyses, whether they must instead be conducted by a trained linguist (native speaker or not) or whether a team approach should be taken (Cambier-Langeveld, 2014; Nolan, 2012).

The findings of the present investigation may be of relevance to this debate as, it may be the case that being interviewed by a native speaker of the language that the applicant claims to be speak, could help the applicant to 'imitate' the accent or dialect via convergence towards the interviewer. Conversely, being interviewed by someone who does not share the same language background or accent, may influence the applicant's speech causing them to deviate from their typical accent or dialect, subsequently resulting in a negative decision. For example, if an asylum seeker claimed to be from Iran and claimed to speak standard Persian it may make a difference whether they were interviewed by a native speaker of standard Persian, versus a speaker of another variety of modern Persian (e.g. Dari Persian or Tajik Persian). Although these different varieties of Persian are said to be largely mutually intelligible (Beeman, 2005), differences may be present between the varieties that could influence the asylum seeker's speech.

Although it is recognised that convergence is unlikely to result in a speaker completely mirroring the accent of their interlocutor, convergence on only a small number of phonetic parameters could still potentially have a bearing on the outcome of a LADO assessment. As previously mentioned in this chapter, a number of scholars have proposed that accommodation may be more likely to occur, and may occur to a stronger degree, in respect of socially salient parameters (Babel, 2009; Cao, 2018; Trudgill, 1981). It may also be the case that these same parameters are the ones that analysts will focus on when determining the accent profile of the speaker and a subsequent decision in respect of their claimed identity. For this reason, the results of the present investigation may have implications for disputed nationality cases as well as for LADO research more broadly.

2.2.3. Development of spoken dialogue systems

A final practical application of speech accommodation research has been to find ways to improve the effectiveness and usability of spoken dialogue systems (SDS). A number of human-to-computer interaction studies have examined the implications of having virtual interlocutors or SDS that accommodate to the human user. Levitan et al. (2016) developed an architecture and algorithm for implementing acoustic-prosodic entrainment in SDS whereby the SDS entrains to their users' speech. In a series of pilot studies, Levitan et al. (2016) examined how accommodating behaviour on the part of an SDS is perceived by the user by testing whether subjects were more likely to ask for advice from a conversational avatar that entrained towards the user versus one that did not. The experiment was conducted using English, Spanish and Slovak SDS. In the English pilot, it was determined that subjects were more likely to ask advice from the entraining avatar. However in the Spanish and Slovak studies, the opposite trend was found whereby subjects preferred the disentraining avatars.

More recently, a related study tested the effects of entrainment in speech rate, pitch and intensity on how participants respond to their virtual interlocutor and found that there was a tendency for participants to perceive disentraining avatars as more competent (Gauder, Reartes, Gálvez, Běnuš, & Gravano, 2018). Another similar study explored how prosodic entrainment correlates to the trust of a human user in an avatar's ability to provide good advice (Beňuš et al., 2018). Their main finding was that females tend to trust the avatar whose average pitch, intensity and speech rate in a turn are locally disentraining (i.e. do not converge towards the user). Beňuš et al. (2018) also noted that applications might need to increase the entrainment effect in order to affect trust of humans towards avatars. The findings of accommodation studies such as these can help to inform SDS developers of how to make modifications in order to improve the effectiveness and user satisfaction of future systems.

2.3. Conclusion

This chapter has provided an overview of what speech accommodation is, why it is thought to occur and how it has been measured in previous empirical research. A summary has also been presented of the practical implications of speech accommodation. A range of different theories relating to the underlying mechanisms involved in speech accommodation, from the perspectives of sociolinguists, psycholinguists and cognitive scientists, were briefly summarised with the main intention being to acknowledge the competing theories and methodologies.

The research presented in this thesis does not specifically aim to provide evidence as to the causes of accommodation, as it is believed that an experiment of this nature would require a much more controlled experimental design, most likely involving non-interactive speech. Instead, the primary focus of the present investigation is to provide an insight into how individual speakers adapt their speech across different forensically-relevant speaking scenarios and whether speech varies across the three West Yorkshire boroughs as a whole. For this reason, a greater emphasis is placed on ecological validity than experimental control.

3. Data

The data analysed in this thesis is from the West Yorkshire Regional English Database (WYRED) which was collected by myself and my colleagues, Dr Erica Gold and Sula Ross, between 2016-2019 (Gold et al., 2018). WYRED is the largest forensically-relevant database of British English speech, containing recordings from 60 speakers from each of three metropolitan boroughs of West Yorkshire: Bradford, Kirklees and Wakefield. In total, 180 participants were recorded undertaking four style-controlled tasks. A primary aim of the WYRED project was to develop a forensically-relevant database and make it publically available for use by researchers, forensic phoneticians, and any other interested parties. A further aim was to explore the generalisability of population data, relating to a range of different speech parameters, in order to identify the level at which speaker groups need to be defined to FSC purposes (Gold et al., 2018, p. 2748).

The research presented in this thesis examines recordings from a subset of 30 WYRED participants. It is important to note that the WYRED database was designed with the overarching research project's goals in mind, as opposed to the specific aims of my PhD research. For this reason, some of the methodological choices that I have made may have been different if I had been eliciting the data for the primary purpose of investigating speech accommodation. Some of the limitations related to this will be discussed in more detail below. It is also necessary to highlight here that my employment as a Research Assistant on the WYRED project was entirely separate from my PhD studies at the University of Huddersfield. While the research presented in this thesis was carried out independently, the collection of the data was a collaborative effort of the whole research team. In terms of my contribution to the WYRED project, I was involved in deciding which boroughs of West Yorkshire would be examined in this project and how the speaking tasks would be formatted. I also played a major part in recruiting participants, facilitating the recording sessions, as well as contributing to the process of transcribing the recordings and organising the files so that they could be made publically available. This chapter provides further details in relation to each of these aspects by outlining the fieldwork methods used in the project, and it also presents a description of the participants analysed in this thesis and the speaking tasks that they each completed. Prior to this, Section 3.1 introduces the research location for this investigation, by providing a brief overview of West Yorkshire's history and geographical location. A summary of the salient accent features associated with Yorkshire English is also provided and motivations for examining local level regional variation across West Yorkshire are set out.

3.1. The community

3.1.1. Location

West Yorkshire is situated in the North of England, within the region of Yorkshire and The Humber. In June 2020, the latest estimated population for West Yorkshire was 2,332,469, which accounts for approximately 42% of the population of the Yorkshire region as a whole (Office for National Statistics, 2020). Figure 3.1 presents a map of Great Britain alongside a detailed map of West Yorkshire which displays the former and modern district boundaries. The district boundaries in place today came into effect in April 1974, under the terms of the Local Government Act 1972 whereby the 53 former local government districts were amalgamated. As can be seen in Figure 3.1, West Yorkshire now consists of five metropolitan boroughs (Bradford, Calderdale, Kirklees, Leeds and Wakefield). It is bordered by the counties of Lancashire, Greater Manchester, Derbyshire as well as North and South Yorkshire.

This investigation examines three of the five West Yorkshire boroughs: Bradford, Kirklees and Wakefield. When comparing the former district boundaries of these three boroughs, it can be seen that Wakefield contained many previously rural districts whereas Kirklees was much more urban. It is possible that the differences in terms of demographic makeup and history of the three boroughs may result in linguistic differences across the region today.

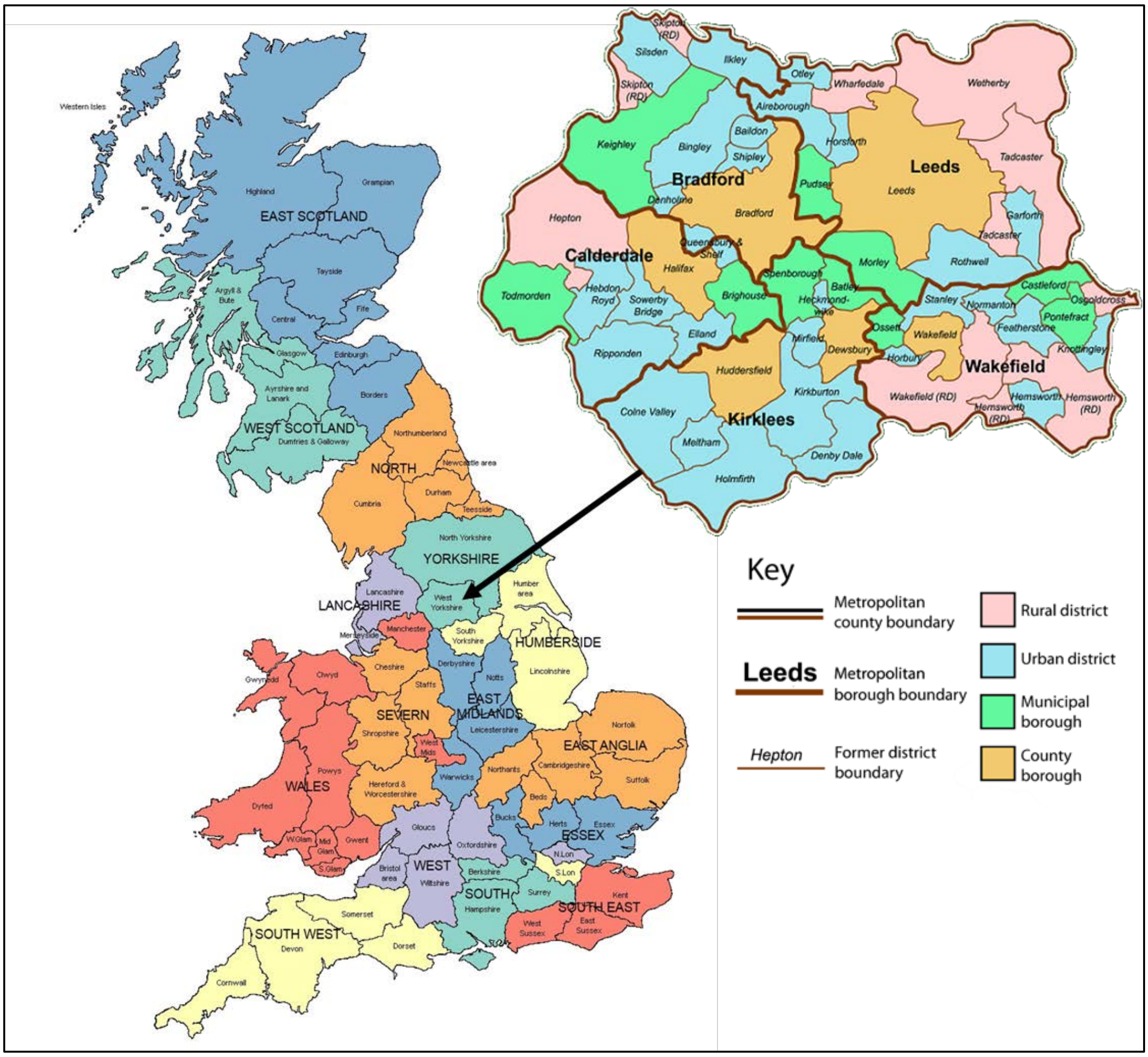


Figure 3.1. Map of Great Britain and West Yorkshire, with the former and modern district boundaries of West Yorkshire marked.

3.1.2. History

In 1888, Wakefield was the first area in West Yorkshire to be granted city status. However, as a consequence of the Industrial Revolution, the areas of Leeds and Bradford grew rapidly and were subsequently also granted city status in 1893 and 1897, respectively. Today, Leeds is the largest city in Yorkshire with an estimated population of 793,139 (Office for National Statistics, 2020). Figure 3.2 illustrates the relative population of each of the boroughs within West Yorkshire based on the ONS population estimates (Office for National Statistics, 2020). Here, it can be seen that the populations of Bradford, Kirklees and Wakefield are relatively similar to one another and that Calderdale has the smallest population of the five boroughs.

The same pattern exists in terms of population density, with Leeds having the largest population density and Calderdale having the smallest.

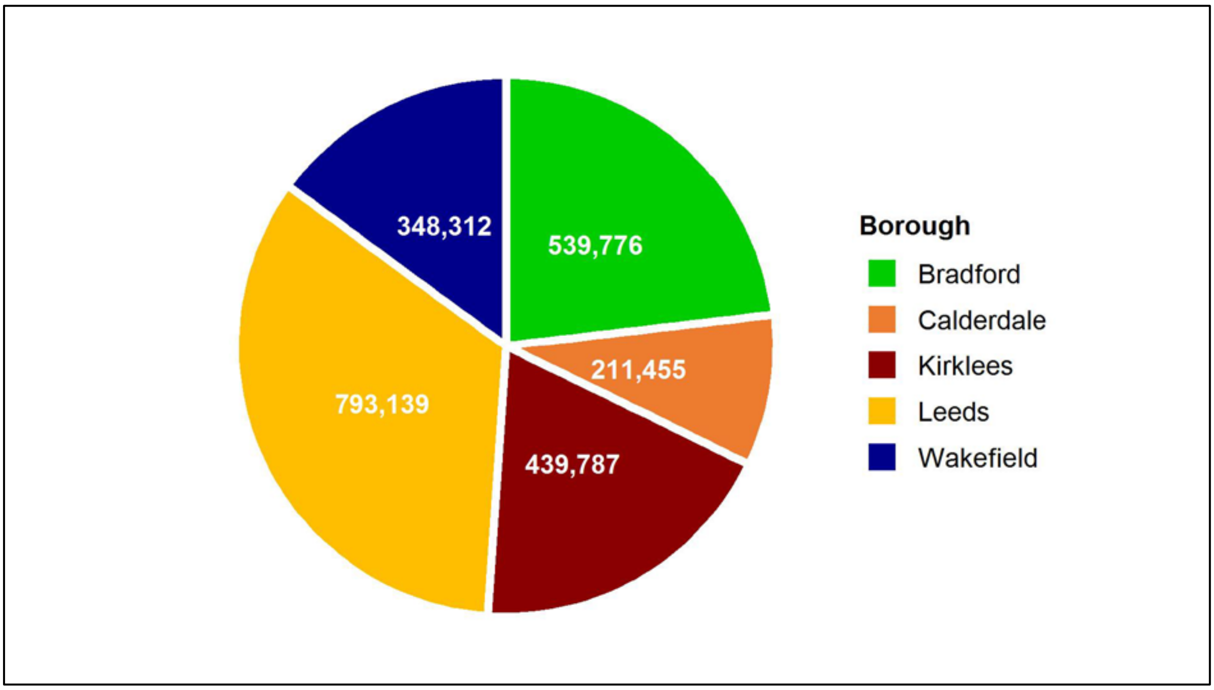


Figure 3.2. Proportional population of the five West Yorkshire boroughs.

The area of West Yorkshire has strong associations with the coal, wool and iron ore industries. Mining was previously a significant employer in the traditional coal mining areas of West Yorkshire, which included Wakefield and South and East Leeds. In addition to coal mining, the textile industry was particularly strong across many parts of West Yorkshire, with Leeds becoming a major mill town during the Industrial Revolution and Bradford gaining a reputation as "the wool capital of the world". However, throughout the 20th century the wool and cloth industries declined heavily. Nevertheless, Leeds is said to be the UK's fastest growing city, with one of the most diverse economies and the third largest jobs total by local authority area in the country (Leeds City Council, 2020). One of the reasons that the WYRED research team chose to examine the three boroughs of Bradford, Kirklees and Wakefield, and to exclude Leeds and Calderdale, was that these three boroughs were similar in terms of population size and demographic makeup. It was also considered that Leeds was more metropolitan and therefore many people living there were likely to be highly socially mobile and high levels of variability were expected across different areas within the city of Leeds.

3.1.3. Demographics of the three boroughs

The present study investigates regional variation, as well as speech accommodation effects, in speakers from the three boroughs of Bradford, Kirklees and Wakefield. Based on data from the 2011 UK census, it is possible to gain an insight into the broad demographics of each of these areas, although it is acknowledged that some changes are likely to have occurred in the decade since the data was collected. Across all three boroughs, the top five industries in which the residents worked were: 'wholesale and retail trade', 'human health and social work activities', 'manufacturing', 'education' and 'construction' (Office for National Statistics, 2012b). The percentage of people employed in each of these sectors was largely similar across boroughs, although Wakefield had higher rates of people working in retail than Bradford and Kirklees, and Kirklees had the highest rates of people working in manufacturing.

With regards to languages spoken across the three boroughs, the vast majority of people reported having English as their main language; with 85.3% reporting this in Bradford, 91.4% in Kirklees and 96.5% in Wakefield (Office for National Statistics, 2013). After English, the next most commonly reported main languages in use across the boroughs were Panjabi, Urdu, Polish, Gujarati and Bengali (with *Bengali* also including Sylheti and Chatgaya). Figure 3.3 presents the percentage of the population of each borough to use each of these languages as their main language and it can be seen that the distributions vary slightly by borough. The most common main language after English spoken in Bradford and Kirklees was Panjabi whereas Polish was the second most common language spoken in Wakefield.

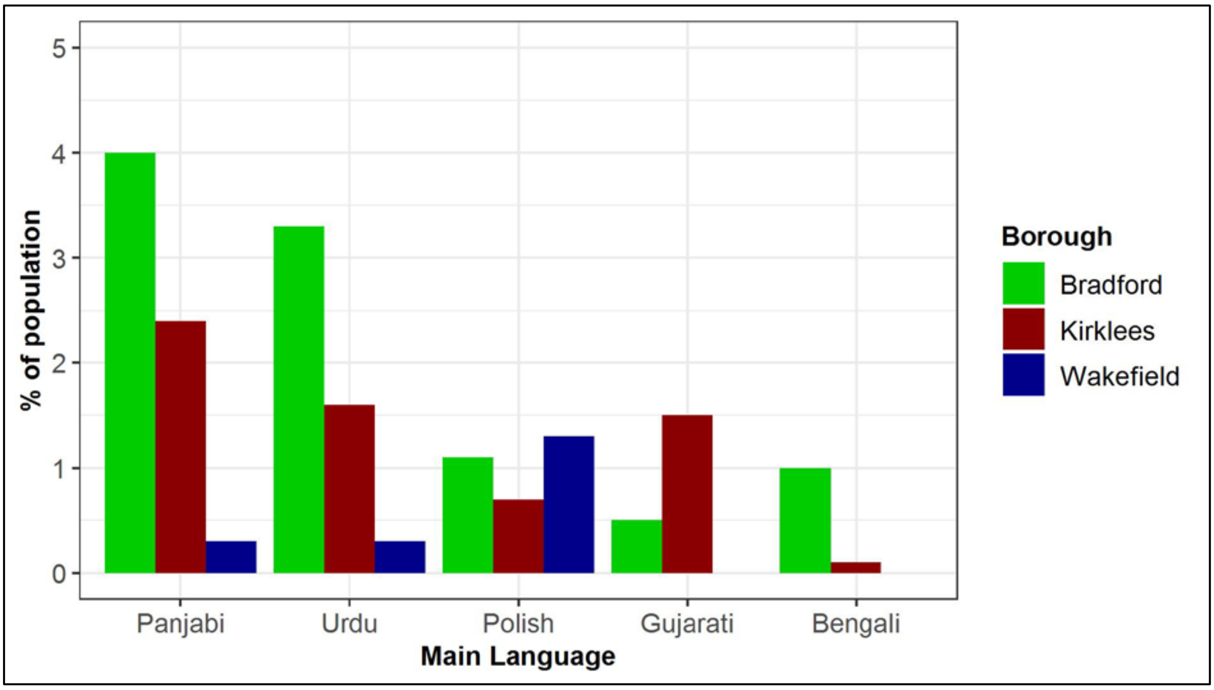


Figure 3.3. Top 5 main spoken languages other than English in Bradford, Kirklees and Wakefield.

In terms of ethnicity, there was also slight variation across the three boroughs. In the 2011 census, the largest ethnic group in the areas of Bradford, Kirklees and Wakefield was White British (at 63.9%, 76.7% and 92.8%, respectively), according to the Office for National Statistics (2012a). After White British, the next most common ethnic groups across West Yorkshire were Pakistani, 'Other White' (i.e. not British or Irish) and Indian. In Bradford, the largest ethnic groups after White British were Pakistani and Other White, who comprised 20.4% and 3% of the population, respectively. In Kirklees, the next largest groups were Pakistani and Indian, who comprised 9.9% and 4.9% of the population, respectively. And finally, in Wakefield the second largest ethnic group was 'Other White' (2.3%) and the third was Pakistani (1.5%). Overall, Bradford appears to be the most diverse of the three boroughs in terms of ethnicity and language use whereas Wakefield appears to be the most homogeneous borough.

In terms of defining the relevant population for FSC casework, the demographic factors listed above would ideally all need to be taken into account and matched to suit the specific case, where possible. For instance, in a FSC involving a questioned speaker that appears to be a young, white, working-class male, the typicality of any similarities between the known and questioned speaker should be assessed in relation to what would be expected within young, white, working-class males more generally. The current investigation deals with a homogenous sample of speakers from the boroughs of Bradford, Kirklees and Wakefield, in that all participants have English as their first and only language, they are all white and are all university educated. This means that the sample examined in this investigation are representative of the majority groups within the West Yorkshire boroughs, but they are by no means representative of the whole population. It is therefore important to be aware that the findings of the case studies in this thesis cannot be generalised to the population as a whole because they relate to a specific demographic group. Furthermore, it should be acknowledged that any sociolinguistic/phonetic differences that may exist across boroughs as a result of differences in terms of language background or ethnicity are unlikely to be observed within the population observed in this investigation.

3.1.4. Overview of Yorkshire English

The present investigation examines how much phonetic variation exists across the West Yorkshire boroughs of Bradford, Kirklees and Wakefield, in respect of two phonological variables: the FACE vowel and intervocalic /t/. Although there are some studies which have examined West Yorkshire speech, it is fair to say that West Yorkshire has received relatively little attention from the sociophonetic community in recent years. Those studies that have focussed exclusively on West Yorkshire speech, have largely investigated one single area as opposed to examining variation across neighbouring areas within West Yorkshire (cf. Easther, 1883; Watt & Tillotson, 2001; Wells, 1982; Whisker-Taylor & Clark, 2019).

The only exception to this that the author is aware of is Petyt's (1985) investigation of the accents of Bradford, Halifax and Huddersfield which indicated that regional variation existed at this very fine-grained level. Specifically, Petyt observed that there were differences between these areas that were "largely of a 'quantitative' nature" rather than it being the case that a particular feature occurred in one area and not in another (1985, p. 356). Most notably, Petyt found that the realisation of /t/ as [ʔ] was significantly more common in Bradford than in Halifax or Huddersfield and that Bradford seemed to be leading the change towards more common usage of this variant. Petyt's findings in relation to /t/ and FACE will be discussed further in subsequent chapters. In recent years, Burland-Gibson (2019) is one of the only investigations which has explored phonetic variation between different areas within Yorkshire. This study examined the areas of Royston, Barnsley and Wakefield and explored the relationship between phonological variation and perceptions of local identity in each of the areas. The findings of this study are discussed in Chapter 4.

Most research involving speakers from West Yorkshire has also included speakers from other areas of Yorkshire and findings have sometimes been reported more generally as examples of "Yorkshire English" (Hughes, Trudgill, & Watt, 2012; Wells, 1982, 1985). Some of the most widely recognised and commonly reported features of Yorkshire English are as follows:

- Monophthongal realisations of FACE and GOAT (Petyt, 1985, p. 162)
- Fronted realisations of GOOSE (Wormald, 2016, p. 196)
- The absence of /h/ word-initially (Hughes et al., 2012, p. 106)
- Definite article reduction (DAR; Jones, 1952)
- The presence of "Yorkshire Assimilation" whereby voiced obstruents become fully devoiced when followed by a voiceless segment (Whisker-Taylor & Clark, 2019)

For a thorough description of a wider range of recognisable features that have been reported in traditional Yorkshire accents, see Wilhelm (2018, sec. 3.1). In recent years, it has been claimed that dialect levelling (i.e. the process whereby regional variants are replaced with either standard or pan-regional ones, leading to differences in regional varieties being reduced) has taken place in parts of the North of England. Consequently, the terms "General Northern English" and "Standard Northern English" have been used to refer to this more broad Northern variety that has emerged. Some of the common features shared by a range of northern varieties include the absence of a FOOT-STRUT split and a TRAP-BATH split (Hughes et al., 2012; Wells, 1982).

In a recent study, Strycharczuk, López-ibáñez, Brown, & Leemann (2020) measured the success of random forest models trained to differentiate a range of Northern English urban accents from one another, based on their full vowel systems, in order to test their hypothesis that many speakers in the North of England are levelling to a pan-regional standard. Results of their investigation revealed a considerable degree of levelling between the accents of Leeds, Manchester and Sheffield. In all three cities they found that happY is typically tense, GOOSE is fronted, FOOL is retracted, and FACE, PRICE and CHOICE are all closing diphthongs. It was noted that all of these features are generally found in Southern British English and therefore provide a sign of dialect levelling toward a "more general British Standard" (Strycharczuk et al., 2020, p. 14). The three cities also shared some more typically Northern features such as a raised STRUT vowel, monophthongal SQUARE and a lack of TRAP-BATH split. Based on this, they argued that the vowel systems for these areas were all representative of pan-regional General Northern English, however, it was noted that some systematic differences existed between them (2020, p. 14). Taking the results of this study into account, it may be the case that any differences that previously existed between the boroughs of Bradford, Kirklees and Wakefield may have reduced as a result of dialect levelling in recent years.

The case studies presented in the present thesis will examine the current state of regional variation in West Yorkshire, in order to explore whether local level variation exists in this area or whether in fact the different accents have levelled to become less distinct from one another. The implications of these findings for forensic speech science will also be considered with respect to the subject of delimiting the relative population for typicality assessments in FSC casework. Specifically, if the case studies in this thesis provide evidence of local level variability between the boroughs of Bradford, Kirklees and Wakefield, this would indicate that reference populations need to be narrowly defined at the borough level. However, if speakers from all three boroughs are found to have similar realisations of the parameters under investigation, this would suggest that it may be appropriate to use a broader reference population defined at the regional level of West Yorkshire.

3.2. Fieldwork

This sub-section describes the fieldwork that was undertaken to collect the WYRED database. Details regarding ethical clearance, recruitment, recording procedures and data management are set out. Information is also provided about the specific 30 participants that were analysed in this investigation as well as the speaking tasks that were considered.

3.2.1. Ethics

Ethical clearance was granted for the WYRED project by the Ethics Committee within the School of Music Humanities and Media at the University of Huddersfield. In line with this, all participants read and signed an information sheet and consent form prior to taking part in the study. All participants in the database were assigned a participant number to be used in place of their names in order to maintain their anonymity. These numbers have been retained in this study. All participants were all compensated for their participation.

3.2.2. Recruitment

A range of strategies were used to recruit participants to take part in the WYRED project. Due to the project being carried out at the University of Huddersfield, most of the in-person recruitment was carried out on campus via distribution of flyers and posters, as well as in-class presentations. The opportunity to take part in the project was also advertised via email and on social media platforms such as Facebook and Twitter. Participants were also encouraged to refer any of their eligible friends or family members to participate. All interested participants were invited to complete an online Expression of Interest form, which enabled the research team to screen them against the project's specific eligibility requirements. To be eligible to participate, participants had to be male, aged 18-30, have grown up and gone to school in either Bradford, Kirklees or Wakefield, have English as their first and only language and have been from an English-only speaking household. Participants also had to confirm that they did not have any speech or hearing impairments. Any interested participants who met these eligibility requirements were invited to book their first recording session via email.

The majority of research in the field of forensic speech science tends to be conducted using male speech, due to the fact that there is a significant lack of reference databases containing female speech available. The TUULS corpus of North Eastern English (Watt, Llamas, French, Braun, & Robertson, 2018) is one of the few corpora to include both male and female speech. In recent years, The International Association for Forensic Phonetics and Acoustics (IAFPA) have attempted to encourage more research into female speech by inviting applications for research grants on this specific topic (IAFPA, 2018). However, it still remains that the overwhelming majority of FSC cases involve the analysis of male voices, with female speech only appearing in approximately 10% of FSC cases in the UK (J.P. French, personal communication, 2020).

WYRED did not include recordings of female speakers as it aimed to be comparable with the existing DyViS database (Nolan et al., 2009) which consists of male, SSBE speech which was collected in Cambridge for the DyViS project. Prior to the collection of WYRED, DyViS was previously the only forensically-relevant database of British English speech in existence and therefore a wealth of studies have collected population data for the field using this database (c.f. Earnshaw, 2014; Gold, 2014; Hughes, Wood, & Foulkes, 2016 *inter alia*). It was intended that by analysing a comparable population of speakers from the North of England, future studies could explore regional differences between these two areas. For this reason, this investigation only examines male speech, from a relatively homogeneous group of speakers.

3.2.3. Recording equipment

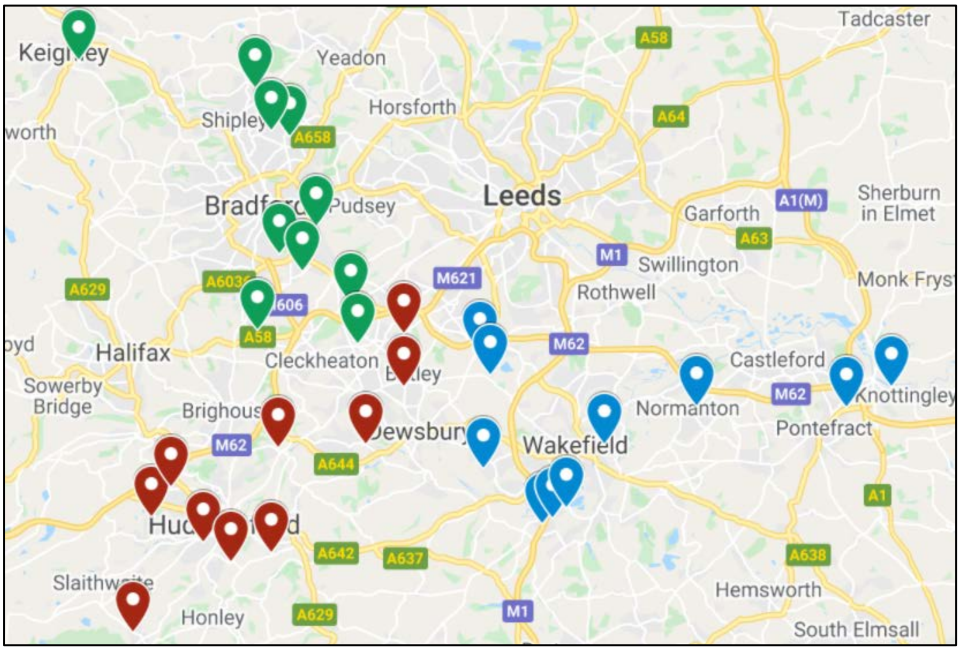
All of the WYRED tasks were recorded using high-quality equipment inside a professional 2.3 by 1.6 meter, purpose-built sound booth at the University of Huddersfield. During all tasks, the participants wore a Sennheiser HSP 4 omnidirectional headband microphone that was placed approximately 2 cm away from their mouth. Recordings were made on a Marantz PMD661 MKII Handheld Sold State Recorder in PCM-WAV format (44.1 kHz, 16 bit). The researchers also wore a Sennheiser HSP 4 omnidirectional headband microphone and their speech was recorded on a separate Marantz PMD661.

3.2.4. Data management

Prior to the WYRED database being made publically available, all audio files were stored on a paid for secure server within the university, with access privileges only granted to the WYRED research team and IT staff. Digital copies of the signed consent forms were also saved on the secure server. Now that the project is complete, in line with the ESRC funding body regulations, the database is stored and maintained by the UK Data Service. All audio files, transcripts and the materials used to carry out each speaking task are available to download from https://reshare.ukdataservice.ac.uk/854354/.

3.2.5. Participants

A subset of 30 participants were selected from WYRED to be used in all four of the case studies presented in this thesis. Participants were equally distributed across the boroughs of Bradford, Kirklees and Wakefield, in order to facilitate an analysis of local level regional variation within West Yorkshire. Participants were classified as being from one of the three boroughs based on the postcode of where they grew up and went to school. Although some of the participants from Bradford and Wakefield had recently moved to Huddersfield whilst studying at the university, all participants had lived in the borough that they were from for the majority of their life. Figure 3.4 presents a map of West Yorkshire with all 30 participants plotted according to where they are from, and colour-coded based on their borough. Throughout this thesis, plots demonstrating differences across boroughs will retain this same colour system with Bradford represented in green, Kirklees in red and Wakefield in blue.



Borough:
- Bradford
- Kirklees
- Wakefield

Figure 3.4. Map of West Yorkshire with participants plotted according to their postcode.

At the time that the 30 participants selected for inclusion in the present investigation were chosen, only approximately 75 of the 180 WYRED participants had been recorded. The selection criteria used to determine which of these participants would be analysed were as follows: (1) their first speaking task had to have a duration of at least 20 minutes, which was the average length of recording for this task, (2) there had to be no evidence of obvious attempts to disguise their West Yorkshire accent, (3) their partner also had to meet these requirements. The reason for specifying a minimum duration for the first speaking task was that it was important to ensure that the participants were talkative and that there was a sufficient amount of speech for analysis. The participants' recordings were also screened to ensure that they were not overly acting or attempting to modify their accent[1] because it was important that their speech was relatively representative of their usual way of speaking. However, it is recognised that the fact that the participants' speech was being recorded in a laboratory setting was likely to result in less natural speech, as a consequence of the Observer's Paradox. This term, coined by Labov, refers to the challenge of finding out how people talk when they are not being systematically observed, when we can only obtain these data by systematic observation (Labov, 1972, p. 209). As participants were recorded alongside another participant for one of the tasks, it was necessary to ensure that both participants in each pair met these requirements. For this reason, the first five pairs of participants from each of the three boroughs considered suitable were included in this study.

The 30 participants included in this study have an average age of 21.8 years and their ages ranged from 19-29 (standard deviation (SD) = 2.37 years). Table 3.1 contains a summary of the participants' ages broken down by borough. All participants had English as their first and only language and were raised in an English-only speaking household. In terms of ethnicity, the majority of the participants who took part in the WYRED project were white and, in terms of educational background, all 30 participants were enrolled on undergraduate or postgraduate degree courses at university or had already completed a university qualification at the time of recording.

[1] One particular participant attempted to portray a character during one of the tasks, by applying an unnatural speaking style and accent that was very different to the way in which he spoke during the other tasks and whilst talking to the researchers when he was not being recorded. As a result of this, he was not included in this study, or in the main WYRED database.

Table 3.1. Summary statistics relating to the age of the 30 participants across boroughs.

Borough	Number of participants	Age (years)		
		Average	SD	Range
Bradford	10	20.8	1.72	19-24
Kirklees	10	22.4	3.01	19-29
Wakefield	10	22.2	1.83	19-25

The availability of a relatively homogeneous population of speakers meant that it was possible to examine regional variation across West Yorkshire, as well as the effects of speech accommodation on the phonetic parameters of interest, without having to account for the influence of social factors such as age, gender and language background. However, for this reason caution must be taken not to overgeneralise when interpreting the findings of the case studies presented in this thesis, as it could be the case that females (or speakers from other demographic groups) behave differently in terms of speech accommodation behaviour and levels of variation across boroughs may also be different across other speaker groups.

It should be noted that all except one of the 30 participants were, or had previously been, a student at the University of Huddersfield. It must therefore be acknowledged that the participants from Bradford and Wakefield had connections to Huddersfield and will most likely have spent time with people from Kirklees during their studies. As a result of this, it is possible that these participants may have already been influenced by long-term accommodation effects towards Kirklees speakers, or indeed other accents that they were exposed to during their time at university. A possible consequence of this could be that levels of variability between the three boroughs may be less extreme within these participants than they would be if comparing groups of speakers who had only ever lived and studied in the borough that they were from. However, as the participants did not move to Huddersfield until they were at least 18 years of age, it is considered unlikely that their accents will have been strongly affected by leaving their home borough.

3.2.6. Questionnaire

Prior to entering the sound booth to undertake the first speaking task, participants were asked to complete a survey, so that various metadata could be collected. The purpose of collecting this information was to enable researchers making use of the WYRED database to be able to take this into account, either when analysing their results or when selecting samples of participants to analyse. Responses to the survey provided information relating to how many years they had spent outside of West Yorkshire, their employment status and job role (if employed), information about where their parents were from, as well as information about their personal sense of regional identity. Specifically, participants were asked to answer which of the following phrases they identified with the most: *British, English, Yorkshire, West Yorkshire, Bradford, Wakefield, Kirklees, Huddersfield, Other*. Although all participants signed up to take part in the project on the basis of being from either Bradford, Kirklees or Wakefield, this supplementary information provided an indication of how meaningful this label was to them. The labels that were selected by the participants examined in this study will be discussed in Chapters 4 and 6, when considering the regional variability findings. Appendix 1 contains metadata that was collected for each of the participants included in this investigation, as part of the WYRED project. For further details about other metadata that was collected, please see Gold et al. (2018).

3.2.7. Speaking tasks

The present study considered three of the four speaking tasks from WYRED. One of the primary aims of WYRED was to be forensically-relevant and therefore the tasks were designed to mirror some of the types of speaking situations that are typically recorded and submitted for FSC work. An important element of this was to try to elicit a range of speaking styles, via tasks involving different interlocutors, as there is often a mismatch in terms of speaking style between the reference and questioned samples, and evidential samples almost always involve different interlocutors (c.f. Foulkes & French, 2012; Nolan, 1991). One of the main reasons for this is that, in the UK at least, known samples often involve more formal interactions with police officers (e.g. recordings of police interviews), whereas questioned samples do not. Another way in which the WYRED tasks were designed to be forensically-relevant was that they elicited both short and long recordings, as FSCs often involve the analysis of relatively short speech samples.

A further element that had to be taken into account when designing the speaking tasks was the fact that FSCs often involve telephone call recordings which are typically bandwidth limited as a result of the telephone transmission. For this reason, one of the tasks involved the participants having a telephone conversation which was recorded in both studio and telephone quality so that the effects of channel mismatch could be tested using the WYRED database. However, it should be made clear that this thesis exclusively uses the studio quality recordings, as it was not the aim of this study to examine how speech parameters were affected by telephone transmission.

Furthermore, it is almost always the case that evidential recordings analysed in forensic casework are recorded on separate occasions, with varying lengths of time between them (Drygajlo et al., 2016; Rose, 2003). In line with this, non-contemporaneous speech was collected by recording the tasks over two separate sessions. The first two tasks were carried out in the first session and the third and fourth tasks were completed in the second session. In the case of the participants included in this study, the median time between recording sessions was 13 days (ranging from 6-42 days; SD = 7.15 days). As the tasks were recorded over two sessions, it is possible that there may have been larger within-speaker differences evident across speaking tasks than there would have been if all tasks had been recorded on the same day. However, no attempt was made to draw conclusions about long-term accommodation across the two recording sessions due to the fact that the time between recordings sessions varied for each individual participant and many other factors could not be controlled, such as who the participants were in contact with between the two recording sessions. Instead, the aims of the present investigation were to consider the extent to which the participants' speech varied as a consequence of short-term accommodation occurring as a result of the combined influence of different interlocutors and different speaking styles across speaking tasks.

It must be acknowledged that it is not possible to collect entirely authentic speech samples in a laboratory setting that can adequately represent the kinds of real-life scenarios that are typically encountered in FSC casework. For instance, examples of very raised and shouted speech that can occur in real life as a consequence of heightened states of emotion cannot easily be replicated under controlled conditions. Furthermore, the varying degree to which participants feel comfortable in a laboratory setting can influence the type of speaking styles that can reliably be elicited. Nevertheless, the speaking situations presented in the WYRED database were considered to be suitable for the purposes of this investigation, as they involved forensically-relevant types of interaction which captured a range of speaking styles and involved different interlocutors.

Table 3.2 presents an overview of the three tasks included in this investigation that each participant was recorded undertaking and provides details of the speech style elicited and the participants' interlocutors. The speech elicited across all tasks was considered to be semi-spontaneous, as it is not read speech but cannot be considered to be truly spontaneous due to the laboratory setting in which it is recorded. All of the WYRED recordings were accompanied by orthographic transcriptions which were produced manually in Praat TextGrids. A decision was taken not to analyse the participants' Task 2 recordings in any of the investigations reported in this thesis. The reason for this was that I played the role of the participants' accomplice in the Task 2 recordings. As some of the investigations included in this thesis involved analysing the speech of the participants' interlocutor, it was deemed inappropriate to include these recordings as I may have subconsciously adjusted my speech as a result of knowing what was to be investigated. For further information regarding the Task 2 recordings, please see Gold et al. (2018).

Table 3.2. Studio quality recordings from WYRED used in this investigation.

Task	**Speech Style**	**Interlocutor**
1: Mock Police Interview	Interactive task; formal	Female researcher from Gateshead (same for all participants)
3: Paired Conversation	Casual conversation; relaxed	Male participant from the same borough (different for all participants)
4: Answer Message	Time-constrained monologue; stressful	No interlocutor

3.2.7.1 Task 1: Mock Police Interview

The experimental design for Task 1 involved a simulated police interview scenario in which participants had to provide certain information whilst denying anything incriminating. This task is an extension of the HCRC map task technique (Anderson et al., 1991) which was modified for the DyViS project (Nolan et al., 2009) and then further adapted for the purpose of the WYRED project (Gold et al., 2018). Using a fictitious map on an iPad as visual stimuli, participants were provided with information enabling them to answer the police interviewer's questions and certain keywords were elicited which included a range of particular phonetic variables. Participants were informed that they should be as co-operative as possible, however, they must avoid disclosing any incriminating information. All information that could be disclosed was shown in black and anything incriminating was shown in red text.

The role of the police interviewer was played by a white, female researcher from Gateshead (a large town on the southern bank of the River Tyne opposite Newcastle upon Tyne, England). The researcher has a self-reported accent of "a Geordie base with 10+ years of Lancaster influence" and therefore her accent was considered to be relatively distinct from the West Yorkshire participants, although still a variety of Northern English. The decision to have this person play the role of the police interviewer was based on her being part of the WYRED research team, rather than her having a specific accent. For the purposes of the accommodation studies reported in this thesis, the researcher's accent variety was considered to be suitably different from the participants' accents for her to be considered a 'community outsider' (i.e. someone the participants would recognise as being from outside the region).

The interviews took place face-to-face, with the participant and the researcher sitting opposite one another with a table in between them. This meant that participants could rely on speech and facial gestures to communicate effectively. All participants were unfamiliar with the researcher and had met her for the first time on the day of the first recording session. When interviewing each of the WYRED participants, the researcher largely asked the participants the same questions and rarely deviated from a predetermined interview structure, except when it was necessary to prompt participants to provide further information or to follow up on the responses they gave. In order to ensure that the same topics were covered in every interview, the researcher could refer to a set of notes throughout the interview, however, every effort was made to avoid producing read speech and for the task to feel like a real police interview as far as possible. It was therefore considered that the Task 1 recordings were all fairly consistent in terms of input from the researcher and the style of the speech elicited. Due to the serious nature of the scenario presented in the task, and the stressful element of having to hide any incriminating information, the speech style elicited from the participants was considered to be relatively formal. Although it is acknowledged that the role-play nature of the task may not necessarily have been taken seriously by all participants, feedback from the participants indicated that they tended to find the task quite challenging and stressful. This was in line with the reported feedback from the DyViS participants in their Task 1 recordings (Nolan et al., 2009, p. 43). It was therefore considered that some resemblance of the type of power dynamics that would usually be at play in real police interviews was achieved.

The average length of the Task 1 recordings used in this study was 25 minutes (ranging between 20-34 minutes; SD = 223 seconds). Although these recordings might be considered relatively short in comparison to some studies that have examined sociolinguistic interviews of up to two hours (e.g. Devlin, 2014), even the recordings of 20 minutes duration were considered to be sufficient for the purposes of the present study. While it is accepted that longer conversations may have presented more opportunity for participants to accommodate over the course of the task, a number of speech accommodation studies have found evidence of short-term accommodation occurring in similar and sometimes shorter timeframes. For example, Schweitzer & Lewandowski (2014) found evidence of convergence in F1 and F2 values occurring during 25 minute dialogues between German speaker pairs. In the pilot phase of Cao's (2018) study, participants were found to accommodate in terms of their vowel formants during map tasks that only lasted for approximately ten minutes. Furthermore, Babel & Bulatov (2012) also found evidence of convergence in f_0 taking place during 30 minute sessions involving three blocks of shadowing tasks which were preceded and followed by pre-/post-task blocks. Based on these previous findings, there was no reason to believe that accommodation could not take place during the Task 1 recordings analysed in this study.

3.2.7.2 Task 3: Casual Paired Conversation

Task 3 consisted of a casual conversation between pairs of participants from the same borough, without a researcher being present. This task was designed to elicit a fairly relaxed speech style that would most closely match natural, everyday conversations. Although the unnatural laboratory setting and the potential influence of the Observer's Paradox meant that it may not have been possible to capture the participants' most informal speech style (c.f. Labov, 1972; Wolfson, 1976), by pairing participants together and limiting the involvement of researchers it was intended that they would feel less self-conscious during this task. A number of previous sociolinguistic studies have paired participants together in this way (Docherty, Foulkes, Milroy, Milroy, & Walshaw, 1997; Stuart-Smith, 1999a; Wormald, 2016). An attempt was made to recruit pairs of friends to take part in the WYRED project so that they could complete this task together, however, it was more common for participants to sign up individually. Consequently, in the majority of cases participants did not know each other and were paired together by the WYRED research team. Wherever possible, participants were paired based on their postcode so that pairs were from geographically-close areas within the same borough. The purpose of this was to try to increase the familiarity between participants.

In the present study, two of the pairs of participants were already friends before they took part in the project, whereas the other thirteen pairs did not know each other prior to participating. This meant that there was an imbalance across recordings in terms of familiarity and therefore it was necessary to discuss the findings of the friendship pairings in detail and to explore whether any signs of familiarity effects were evident. However, it was not possible to examine the influence of familiarity on the findings in a robust way due to the limited number of friendship pairings. Ideally, levels of familiarity would have been taken into account when selecting participants for inclusion in the study, had more friendship pairings been available at the time the participants were chosen.

In this task, participants were provided with topic cards as prompts, although they were instructed that they could discuss any topics they like. The topics cards were adapted from Wormald (2016) with permission, and included the following themes: Hometown, Family & Friends, Work, Education, Language, Travel, Sport, and Hobbies & Interests. The specific topics that were discussed varied across the 15 participant pairings and some pairings relied on the prompts more than others. Similar to Task 1, this task took place face-to-face, with the participant and their partner sitting opposite each other with a table in between them. Once the researcher had provided the instructions and started the recordings, they left the sound booth so that the participants were alone and then returned after approximately 20 minutes to stop the recordings. The average length of the Task 3 recordings used in this study was 21 minutes (ranging between 19-22 minutes; SD = 48 seconds).

3.2.7.3 Task 4: Answer Message

Task 4 related to the fictional crime scenario from the Task 1 police interview and involved participants leaving an answer message in a time-pressured situation. Participants were instructed to contact their fictional brother and ask him to hide or dispose of any incriminating evidence. They were provided with some brief examples of evidence to discuss, which included some of the keywords mentioned in Task 1, but they were encouraged to talk about additional unprompted information. They were instructed that they had to convey as much information as possible within three minutes but if they ran out of things to say before this time, they could terminate the call early.

As the participants' task was to leave an answer message, they spoke into a telephone and their speech was recorded over a cordless BT Diverse 7410 Plus landline telephone onto a Tiptel 540 answer machine. During this task, participants were left alone in the sound booth without a researcher present. Although the answer message was recorded over multiple channels, this thesis exclusively uses the studio quality recordings. This meant that it was not necessary to account for any technical telephone effects that have previously been reported in telephone quality recordings due to the bandpass filtering effect, such as raised F1 vowel measurements or an increased impression of denasality (Byrne & Foulkes, 2004; Künzel, 2001; Stevens & French, 2012). The average length of the Task 4 recordings was 2 minutes (ranging from 90 seconds to 3 minutes; SD = 29 seconds). Although these recordings were extremely short, they were considered to be forensically relevant both in terms of the limited duration but also because answer messages are often submitted as questioned samples for FSC work.

3.3. Analytical approach

The decision to examine the FACE vowel and intervocalic /t/ in this thesis was based on a number of factors. Firstly, as I was interested in examining local level variation across the three West Yorkshire boroughs, parameters were selected that were predicted to be variable across the region. Previous sociolinguistic studies of West Yorkshire English had indicated that FACE has relatively high levels of variation across the different boroughs of West Yorkshire in comparison to many other vowel sets (Burland-Gibson, 2019; Easther, 1883; Hughes et al., 2012; Petyt, 1985), and productions of /t/ had also been shown to vary across the region (Petyt, 1985). Secondly, I wanted to explore how both regional variation and speech accommodation were influenced by social salience and therefore parameters were chosen that were both considered to be salient within the speech community under investigation, but that also varied in terms of levels of awareness. Specifically, realisations of /t/ as [ʔ] were considered to be highly salient in that T-glottaling is generally something that speakers are aware of in their own speech and in that of others around them (Alderton, 2020, p. 43). The FACE vowel was also considered to be socially salient within West Yorkshire, albeit to a lesser degree than T-glottaling. A more in-depth discussion of salience is provided in later chapters of this thesis. Another motivation for examining FACE and /t/ was that they each represent a type of parameter that is commonly examined in FSC casework, which can be analysed using different methodological approaches.

In a survey of international practices in FSC casework, Gold & French (2011) found that all respondents reported analysing vowel and consonant sounds in the course of their FSC examinations. With regards to the specific analytical approaches taken by these experts, it was reported that 97% of experts examined formants when analysing vowels and of those undertaking formant examinations, it was reported that “all measure the second resonance (F2); 87% of respondents reported measuring F1 and an equal percentage reported measuring F3” (2011, p. 300). In respect of consonants, 88% of respondents reported evaluating auditory quality. Furthermore, plosives were reported to be analysed relatively frequently in FSC cases in comparison to consonants with other manners of articulation, with only fricatives being examined more frequently, on average (2011, p. 301). Based on the reported conventions within the field, it was considered that FACE and /t/ were both features that would be routinely examined as part of a FSC case. The analytical approaches taken in this thesis were also deemed to be appropriate in the context of FSC casework. In this thesis, FACE is analysed acoustically by measurement of vowel formants (F1~F3), whereas intervocalic /t/ is analysed using an auditory approach.

For each parameter, an assessment of regional variation across the three boroughs is presented, followed by a subsequent examination of within-speaker variability across the individual speaking tasks to explore the influence of speech accommodation. As the specific methods of analysis were distinct across each of the four case studies, full details of the analytical approach taken for each study are set out across Chapters 4-7 separately. Specific research questions and hypotheses for each of these case studies are also outlined in the subsequent chapters. In the case studies presented in the following four chapters, the recordings of all 30 West Yorkshire participants completing the three WYRED speaking tasks are analysed. In addition, the speech of the researcher in Task 1 was also analysed as part of the accommodation studies presented in Chapters 5 and 7. Orthographic transcriptions were not provided for the speech of the researchers in the WYRED database and therefore prior to analysing the researcher's speech, each of the Task 1 recordings were transcribed into Praat TextGrids using the researcher's studio recordings. This task was performed by an undergraduate research assistant at the University of Huddersfield and checked by myself.

Although the WYRED data lends itself well to the investigations presented in this thesis, as it includes forensically-relevant semi-spontaneous speech elicited in paired and unpaired tasks; the order in which the tasks were completed makes the set-up of the speech accommodation studies slightly unconventional. In both non-interactive and conversational accommodation studies, typically an assessment of each participant's baseline is established during a pre-exposure task and then this baseline is compared to the participant's speech during a paired/speech-shadowing task and in some cases a post-exposure task (e.g. Babel, 2009; Pardo, 2006; Staum Casasanto, Jasmin, & Casasanto, 2010). However, in this investigation the participants were not required to complete a pre-exposure task prior to completing the two paired tasks. Task 1 was the first task completed during the first recording session, whereas Tasks 3 and 4 were completed consecutively during the second session. This meant that there was no "baseline" speech available, in the conventional sense, and therefore speech from the

Task 4 recordings was treated as being an example of each participant's "baseline" instead. This could be considered problematic as it is possible that any speech accommodation that took place during Task 3, could potentially have persisted into the Task 4 recording. However, it could be argued that any form of completely unbiased "baseline" is impossible to achieve, as even in a pre-task setting it is possible that a participant's production could be influenced by a prior experience, whether that be talking to the researcher before the experiment begins or talking to someone else earlier in the day. Furthermore, due to the very different contexts of each of the three tasks, it is considered that all tasks can be treated as being somewhat independent of one another.

The aims of the accommodation case studies presented in this thesis were to examine how each participant produced FACE and intervocalic /t/ across the three tasks, in order to establish how much these parameters varied within a participant and whether they appeared to be influenced by speech accommodation during the paired tasks. It was intended that the speech from Task 4 would be suitable for building a picture of each participant's baseline speech because they were talking in a monologue-like style, without any interaction or feedback from another person. However, it is not known whether any of the participants had a brother in real life that they were pretending to address, or whether they were imagining a portrayal of similar criminal behaviour on television, and therefore it must be acknowledged there may have been some influence of audience design at play for these participants. It must also be recognised that due to Task 4 being a relatively short task, there were only a limited number of FACE and word-medial, intervocalic /t/ tokens produced by each participant during this task. For this reason, and the fact that participants were speaking into a telephone, in a potentially stressful time-pressured situation, it is possible that the speech recorded during this task may not be entirely representative of each participant's natural, spontaneous speech. For this reason, the term "baseline" is employed in future chapters more as a way of distinguishing speech in this task from speech produced during an interactive paired conversation (Tasks 1 and 3).

In Task 4, it is possible that the participants may have spoken with a slightly faster articulation rate than usual and their speech may have been influenced by the 'Lombard Effect' whereby speakers increase their amplitude when using the telephone, often resulting in an "epiphenomenal effect of raising the speaker's f_0" (Foulkes & French, 2012, p. 567). It must therefore be taken into account that some differences between FACE and /t/ in Task 4 versus in Task 1 and Task 3 may be due to these speaker effects of talking into a telephone. Nevertheless, the speech samples elicited via this task were considered to be suitable for the purpose of this investigation, given the forensic context of the research.

4. Variation in the FACE vowel across West Yorkshire

4.1. Introduction

This chapter presents an investigation into how the vowel in words of the FACE lexical set, as defined by Wells (1982), is realised across West Yorkshire[2]. This particular phonetic parameter is of interest as a wide range of FACE variants have been reported to be in use across the UK, and previous linguistic research into West Yorkshire specifically, has indicated that FACE has high levels of variation across its various boroughs (Burland-Gibson, 2019; Easther, 1883; Hughes et al., 2012; Petyt, 1985). The aim of this analysis is to establish how speakers from West Yorkshire pronounce this particular vowel and also to explore the extent to which speech production varies on a local level across the boroughs of Bradford, Kirklees and Wakefield. If FACE is still found to be highly variable across West Yorkshire, this would provide motivation to treat neighbouring boroughs, such as these, as separate speech communities when conducting socio-linguistic/phonetic studies and forensic phonetic research.

This chapter is divided into 7 sections. The remainder of this section provides background information relating to the FACE vowel and how it has previously been shown to vary according to region, social factors and phonetic environment. In Section 4.2, the research questions are set out and the overall aims of this investigation are explained. Section 4.3 briefly describes the data used for this study and Section 4.4 outlines the methodology used for this investigation. Section 4.5 gives details of how FACE is realised across West Yorkshire and also presents results of how the vowel varies across the region. In Section 4.6, the results of this study are discussed and their implications are situated in a broader context, before Section 4.7 concludes the chapter.

4.1.1. Regional variation

Descriptions based on auditory analyses of the FACE vowel have been documented in many varieties of English from across the UK. Hughes et al. have reported a general pattern whereby we find "wide diphthongs in the south of England, narrow diphthongs further north, and

[2] Some elements of the investigation presented in this chapter have previously been published in Earnshaw & Gold (2018).

monophthongs in northern Lancashire and Yorkshire" (2012: 154). An extensive literature review confirms that this general pattern is largely accurate, with a few minor exceptions. For instance, monophthongal variants are reported in some areas outside of northern Lancashire and Yorkshire. Tables 4.1 and 4.2 present a summary of the precise quality of both diphthongal and monophthongal FACE variants, respectively, that have been reported in various regions across the country. In both tables, the FACE variants have been listed next to the regions in which they've been observed. It should be noted here that all of the findings in relation to FACE presented below are based exclusively on auditory analysis, except for those of Burland-Gibson (2019), Devlin (2014) and Wormald (2016) which also involved an acoustic analysis of the FACE vowel.

Table 4.1. Regional distribution of diphthongal FACE variants.

FACE variant	**Region**	**Reference**
[æɪ]	'Cockney' London English	Hughes et al., 2012, p. 77
	West Midlands	Hughes et al., 2012, p. 98
[æi]	Norwich	Trudgill, 1999, p.125
	Sandwell	Mathisen, 1999, p. 108
[ɛɪ]	Bristol	Hughes et al., 2012, p. 87
	Derby	Docherty & Foulkes, 1999, p. 48
	Leeds	Wells, 1982, p. 364
	Southampton	Hughes et al., 2012, p. 91
	West Wirral	Newbrook, 1999, p. 94
	West Yorkshire (Bradford, Halifax & Huddersfield)	Petyt, 1985, p. 120
[ɛi]	Leicester	Hughes et al., 2012, p. 101 Wormald, 2016, p. 175
	Milton Keynes	Williams & Kerswill, 1999, p. 143
	Reading	Williams & Kerswill, 1999, p. 145
	Royston	Burland-Gibson, 2019, p. 190
[e̞ɪ]	Liverpool	Hughes et al., 2012, p. 113
	Received Pronunciation (RP)	Hughes et al., 2012, p. 52

	South East London	Tollfree, 1999, p. 165
[ëɪ]	West Wirral	Newbrook, 1999, p. 94
[ei]	Cardiff	Mees & Collins, 1999, p. 187
	Leicester	Wormald, 2016, p. 175
	Manchester	Hughes et al., 2012, p. 117
	Royston	Burland-Gibson, 2019, p. 190
[ɪə]	Tyneside	Hughes et al., 2012, p. 154
	East Durham	Devlin, 2014, p. 147

Table 4.2. Regional distribution of monophthongal FACE variants.

FACE variant	Region	Reference
[eː]	Barnsley	Burland-Gibson, 2019, p. 190
	Bradford	Wormald, 2016, p. 175
	Devon	Hughes et al., 2012, p.147
	East Durham	Devlin, 2014, p. 147
	Leeds	Wells, 1982, p. 364
	Mainstream & fashionable Dublin English	Hickey, 1999, p. 275 Hughes et al., 2012, p. 142
	Newcastle	Watt & Milroy, 1999, p. 27
	Sheffield	Stoddart, Upton, & Widdowson, 1999, p. 73
	Wakefield	Burland-Gibson, 2019, p. 190
	West Yorkshire (Bradford, Halifax & Huddersfield)	Petyt, 1985, p. 120
[ɛː]	Bradford	Hughes et al., 2012, p. 105 Wormald, 2016, p. 175
	Hull	Hughes et al., 2012, p. 108
	Lancashire	Hughes et al., 2012, p. 150
	Leicester	Wormald, 2016, p. 175
	Local Dublin English	Hickey, 1999, p. 275 Hughes et al., 2012, p. 142

	Middlesbrough	Hughes et al., 2012, p. 120
[e]	Belfast (London) Derry English	McCafferty, 1999, p. 247
	Edinburgh	Chirrey, 1999, p. 225
	Glasgow	Stuart-Smith, 1999, p. 206
	Yorkshire	Rogers, 2000, p. 113

In Table 4.1, it can be seen that the wide diphthongal FACE variants are largely found in the South and in the Midlands, whereas the narrower variants are observed further north. However, there are some exceptions whereby narrow diphthongs are reported to be in use in the South. For instance, [e̞ɪ] is noted in South East London and in the prestige RP form, which is typically associated with speakers of Southern British English but is also used in other parts of the UK. In Table 4.2, the long monophthongs, [eː], and [ɛː], and the short vowel [e] are reported as being typical realisations of the FACE vowel, and the areas in which these variants have been observed are listed.

A recent study conducted by Leemann, Blaxter, Britain, & Earnshaw (2019) compared reported usage of FACE variants across England, from the English Dialects App Corpus (EDAC; Leemann, Kolly, & Britain, 2018), with findings from the Survey of English Dialects (Orton & Dieth, 1962). The results of this study suggested that dialect levelling may have taken place across many parts of England, with the dominant form in the south and east of England [ɛɪ], gaining considerable ground. However, respondents were only asked to report how they produce FACE in the specific phonetic environment of the word *bacon* and therefore this may have slightly under-represented the levels of linguistic diversity still in existence across England. Strycharczuk et al. (2020) also made use of data collected via the EDAC to examine variation across the cities of Liverpool, Manchester, Leeds, Sheffield and Newcastle. In this study, crowdsourced recordings of the passage “The Boy Who Cried Wolf” were analysed, with each of the English vowels being represented by a single word. In the case of FACE, productions of the word *safety* were analysed. It was reported that FACE was realised as a closing diphthong across all cities, with more diphthongal variants being observed in Liverpool and Manchester. In line with this finding, Baranowski & Turton (2015) reported that the long mid FACE vowel is an upgliding diphthong that glides towards the high front area of the vowel space in Manchester (2015, p. 3).

Perhaps in contrast to the theory that dialect levelling is leading to there being less variation in FACE productions, both Devlin (2014) and Burland-Gibson (2019) found evidence of regional differences in FACE at a local level. Devlin examined FACE within speakers from four villages in East Durham: Dawdon, Easington, Horden and Blackhall and found that while all locations showed a preference for the traditional North East variant [ɪə], the usage rates of this variant in proportion to the variants [eː], [ɛː] and [eɪ] varied across villages (2014, p. 147). In all areas, [eː] was the second most common variant, however, while Blackhall speakers only showed a subtle difference in percentage usage of [ɪə] and [eː], speakers from the other three areas showed marked differences. It was also observed that usage of the minority variants [ɛː] and [eɪ] was higher in the two southern villages, Horden and Blackhall, than in the more northern villages of Dawdon and Easington. Similarly, in Burland-Gibson's (2019) comparison of FACE across the Yorkshire speech communities of Royston, Barnsley and Wakefield, it was found that diphthongal forms of FACE were dominant in Royston whereas long monophthongal variants were dominant in Barnsley and Wakefield (2019, p. 194).

With regards to West Yorkshire specifically, Wells reported the variants [eː] and [ɛɪ] to be in use in Leeds (1982, p. 364), while Rogers reported [e] for the West Yorkshire accent (2000, p. 113). Rogers simply states that “a West Yorkshire accent is presented as an example of northern English” but it is not clear which specific parts of West Yorkshire were included in the study (2000: 113). In line with the findings of Wells and Rogers, other existing phonetic literature has found [ɛɪ] and [eː] to be common variants of FACE in the Bradford and Kirklees accents (Petyt, 1985, p. 120); with [eː] being the most popular variant, described as a long monophthong somewhere between Cardinal Vowel 2 and 3 (Petyt, 1985, p. 162). In more recent years, Bradford English has been said to typically contain an open-mid monophthong [ɛː] (Hughes et al., 2012, p. 105; Wormald, 2016, p. 175). Whereas in Wakefield, FACE has been found to be most commonly realised as [eː], with [ɛɪ] and [eɪ] also being present in this area (Burland-Gibson, 2019, p. 190).

4.1.2. Social variation

It is well recognised within the fields of socio-linguistics/phonetics that linguistic variation is not only conditioned by where a speaker is from (regional variation) but it is also often correlated with social factors relating to the speaker (social variation). For example, factors such as a speaker's age, gender, social class and language background can often influence the way in which a speaker uses language to communicate. These social factors can affect a range of linguistic features such as pronunciation, grammatical, syntactic and lexical choices. As well as social differences, biological differences between men and women can also influence speech, for instance the differing length of vocal tracts tends to result in females having a higher baseline f_0 compared to males. This section describes the findings of studies which have reported ways in which the length and quality of the FACE vowel appear vary depending on social factors including age, gender, social class, and language background.

Burland-Gibson reported that, despite FACE being most commonly realised as [e̞ː] by most speakers in Wakefield, the older females were the only group of speakers who used the diphthongal FACE variant, [ei], in addition to [eː] and [ɛɪ] (2019, p. 179). Wormald found a similar trend in Bradford English whereby younger speakers exhibited less variation across the FACE vowel trajectory than older speakers (2016, p. 175). In contrast to these findings, Mathisen found that in Sandwell, West Midlands, younger speakers used [æi] for FACE whereas older speakers used a slightly narrower [ɛi] variant (1999, p. 108). Additionally, in Wilhelm's study of North West Yorkshire (defined as "Leeds and a mainly rural area comprising the former West Riding county and a small area a few miles north of the Yorkshire Dales"), monophthongal variants for FACE were commonly used by the large majority of speakers aged over 50 (2018, p.28). However, it was observed that speakers within the 9-15 age group tended to use diphthongal realisations (Wilhelm, 2018, p. 12). It should be noted that in this study, FACE was described as either being monophthongal or diphthongal and no transcriptions detailing the specific quality of the vowel realisations was provided. For this reason, the findings of this study are not incorporated into Tables 4.1 and 4.2.

Haddican, Foulkes, Hughes, & Richards (2013) also reported real-time evidence of diphthongisation of the traditionally monophthongal FACE vowel in York. In their study, data collected in 1998 and 2008 were compared and it was found that there was greater acoustic movement from FACE vowel onset to offset in the 2008 sample than in the older sample. In addition to evidence of change over time, it was also observed that the move towards more diphthongal realisations of FACE seemed to be related to the speakers' identification with the local community (2013, p. 387). Overall it was found that diphthongal FACE variants were most prevalent within speakers who did not identify strongly with the community, whereas speakers who expressed the strongest allegiance to the community tended to use the monophthongal variants (2013, p. 396). The authors noted that the relatively slow rate of change from monophthongal to diphthongal FACE may be linked to strong links that community members make between these variants and different local categorisations. This theory was based on the fact that participants frequently identified monophthongal realisations of FACE/GOAT as "typifying York or Yorkshire dialects" (2013, p. 396). In Devlin's (2014) apparent time study in East Durham, an interaction between location and age group was observed, with statistically significant differences in FACE variant distribution in the younger and older speakers in Horden and Blackhall, but very little change over time in the two more northern villages. In all locations, older speakers tended to favour the traditional, local variant [ɪə], and in all cases they had a higher percentage usage of this variant than their younger counterparts (2014, p. 151).

With regards to gender, FACE has been shown to vary across different speaker groups in various areas of the UK. For example, in Liverpool, it was observed that usage of the main local forms [ëɪ] or [ɛɪ] differed according to gender; whereby males tended to prefer the older local form [ëɪ] while females preferred what was originally seen to be a "hyper-correct" variant [ɛɪ] (Newbrook, 1999, p. 96). In Newcastle, [eː], described as the "the unmarked northern mainstream monophthongal FACE variant", was said to be used at high frequencies by women, young speakers and middle class speakers (Watt & Milroy, 1999, p. 40), although a Southern/Midland-type closing diphthong [eɪ] was becoming more common among these speaker groups (1999, p. 28). In contrast to this, [ɪə] was very common among male speakers (Watt & Milroy, 1999, p. 28).

In terms of social class, Williams & Kerswill noted that although [ɛi] was the most popular variant in Milton Keynes and Reading, middle class speakers were said to be more likely to use the RP variant [e̞ɪ] (1999, p. 143). Language background and the speech community a person is from has also been found to affect FACE vowel productions. Wormald (2016) investigated the linguistic patterns of Panjabi English speakers in Bradford and Leicester and compared a range of parameters in their speech with those of Anglo English speakers from the same communities. With regards to the FACE vowel, it was found that both the Anglo and Panjabi English speakers from Bradford retained monophthongal realisations, although the Panjabi English speakers had close, front variants similar to [eː], whereas the Anglo English speakers had more open [ɛː] variants (2016, p. 175). In Leicester, all speakers had diphthongal FACE realisations but the Panjabi English speakers had shorter trajectory lengths with variants such as [ei] being common, whereas the Anglo English speakers had more [ɛi]-like qualities (2016, p. 175). These findings were in line with the previous finding of Stuart-Smith, Timmins, & Alam (2011) where Asian speakers from Glasgow had been found to have closer and more front FACE productions than non-Asian speakers from this area (2011, p. 11). Sharma (2011) also described a difference in FACE realisations between Asian and non-Asian speakers from Southall in London. Sharma reported that in this community the monophthongal FACE variant [e] was considered to be an Indian feature which differed from the diphthongal [eɪ] realisation of Anglo speakers (2011, p. 470).

4.1.3. Variation across phonetic environments

In addition to regional and social factors, when analysing any linguistic feature, it is always important to take into consideration how the phonological environment in which the feature occurs may influence the feature under examination. In order to do this, an understanding of any phonological constraints that exist within the language variety in question is required. For instance, when analysing vowels, it is usually appropriate to consider how adjacent sounds may influence the quality of the vowel, especially during the transition phases in and out of the vowel. Vowel length can also be affected by the position within a word where it occurs (i.e. whether or not it occurs in a stressed syllable). For this reason, it was necessary to consider the results of previous investigations which had discussed how the FACE vowel was realised across a range of phonetic environments and take these findings into account when planning the experimental design for this investigation.

Not unsurprisingly, the phonetic environment does appear to play a role in conditioning the way in which the FACE vowel is realised. For instance, in a number of regions where a monophthongal variant is the most common form, a diphthong can be found in words with <eigh> spellings, e.g. *weight*. In Bradford, Hull and Lancashire FACE is typically realised as [ɛː]; however, in words with <eigh> spellings [ɛɪ] is often used (Hughes et al., 2012, pp. 105, 108, 150). Similarly, Rogers reports [ej] to be in use in this context throughout West Yorkshire (2000, p. 113). In Sheffield [ɛɪ] is also used in this context as opposed to the popular monophthongal variant [eː] (Stoddart et al., 1999, pp. 73–74). Similarly, in Newcastle [æi] may occasionally be heard in these words rather than the more common variant [eː] (Watt & Milroy, 1999, p. 28).

Furthermore, there is evidence of a lexical effect in the specific words *make* and *take* which has resulted in interesting local forms in many places. Specifically, these words have been reported to be realised with the short monophthongal [ɛ] FACE variant in Bradford, Hull, Manchester and Lancashire English (Hughes et al., 2012, pp. 105, 108, 150), as well as in Derby (Docherty & Foulkes, 1999, p. 48), Sheffield (Stoddart et al., 1999, p. 73), Sandwell (Mathisen, 1999, p. 108), and Western Fenland (Britain, 2014, p. 37). Similarly, the open front monophthong [a] was reported for these specific words in Huddersfield English in the 19th century (Easther, 1883, pp. 1, 6). In recent years, the specific pronunciation of the word *take* as [tɛk] appears to have become enregistered and has been used in local advertisements within West Yorkshire to signal a stereotypical West Yorkshire identity. For instance, the First Group's First Bus App advertisements include the slogan "Want easy travel? Tech the bus" which is built on the premise that the words *tech* and *take* will be homophones in West Yorkshire (First Bus North, 2020).

4.2. Research questions and hypotheses

The specific research questions being addressed in this investigation are as follows:

1. How is the FACE vowel realised across the metropolitan boroughs of Bradford, Kirklees and Wakefield?
2. Are there acoustic differences between FACE realisations across the three West Yorkshire boroughs in question and if so, how extreme are these differences?

Based on the previous findings of sociolinguistic research involving vowels in West Yorkshire, it is hypothesised that the FACE vowel will be realised as either a long monophthongal variant or a narrow diphthong and it seems likely that productions will vary in certain contexts, such as in words with <eigh> spellings and the specific words *make* and *take*. For this reason, phonetic environment is taken into account when considering both how FACE is produced and also when considering if there are noticeable differences between the boroughs of Bradford, Kirklees and Wakefield.

With regards to the second research question, previous literature has indicated that the FACE vowel is highly variable across the region and therefore it is hypothesised that there will be distinctions across the three areas. However, it must be noted that the vast majority of previous studies of West Yorkshire have presented auditory representations of FACE and few have substantiated these representations with acoustic measurements. This study is the first to use an acoustic analysis of vowel formants to investigate whether the production of FACE varies across boroughs of West Yorkshire and if so, to what extent. In the present study, a closely defined population of participants are analysed and therefore confounding factors such as age, gender and socio-economic background are largely controlled; making it possible to test the role of different areas within West Yorkshire as an independent factor. However, in drawing conclusions about regional variation across West Yorkshire, it must be acknowledged that social factors are likely to interact with the effect of location and therefore it is accepted that the findings of this investigation may only be applicable to this particular social demographic.

4.3. Data

This chapter investigates local level regional variation across West Yorkshire by analysing the speech of 10 young, male speakers from each of the boroughs of Bradford, Kirklees and Wakefield. Speech was analysed from recordings of each participant undertaking three of the WYRED speaking tasks: Task 1, 3 and 4. Task 1 involved a mock police interview where the participants interacted with a researcher who was in the role of a police interviewer. Task 3 involved participants having a casual paired conversation with another participant from the same borough as themselves. In Task 4, participants were required to leave an answer message for their fictional brother in relation to the crime scenario from Task 1. Table 4.3 summarises the average recording lengths across tasks for participants from Bradford, Kirklees and Wakefield. The recording durations were broadly similar across participants from each of the three boroughs, with Task 1 being the longest task and Task 4 being the shortest. Full details about each of the speaking tasks and further biographical information about the participants can be found in Chapter 3.

Table 4.3. Average lengths of Tasks 1, 3 and 4 across boroughs.

Task	**Average recording length (mm:ss)**		
	Bradford	**Kirklees**	**Wakefield**
Task 1	24:25	24:45	26:45
Task 3	21:02	20:52	21:07
Task 4	02:04	02:18	02:18

4.4. Methodology

This section outlines the methods employed to investigate how FACE is realised across West Yorkshire. It describes how the FACE tokens were selected, segmented and measured, as well as how the statistical analyses were performed.

4.4.1. Token selection

For each participant, a maximum of 35 tokens were manually segmented from each of the sound files of WYRED Tasks 1, 3 and 4. Tokens of FACE were selected from clearly articulated speech where there was no uncertainty as to what the intended target was. Any tokens produced in overlap or when the participant was laughing were disregarded. Tokens were also excluded from the analysis if they were almost fully elided due to co-articulation. Tokens were only selected from mono- or bi-syllabic content words that contained FACE in the stressed syllable position. Care was also taken to select tokens from a wide range of lexical items, phonetic environments, and in cases where multiple tokens of the same lexical item were available, only the first three instances were included. Due to the experimental design of Tasks 1 and 4, a range of specific keywords related to the tasks occurred frequently in almost all of the participants' recordings, such as *Rachel, cables,* and *steak house*. Appendix 2 contains a frequency table with all of the FACE tokens that were produced by the 30 participants.

4.4.2. FACE segmentation and extraction

Tokens were manually segmented in Praat (Boersma & Weenink, 2019) and the lexical item from which the vowel token was extracted was labelled in a TextGrid on the 'FACE' tier. Figure 4.1 illustrates how the FACE token from the word *maybe,* produced by participant #006, was extracted and labelled. For each token, the beginning of the first complete cycle and the end of the last complete cycle of the vowel were marked at zero crossings (see segmentation boundaries B1 and B2 in Figure 4.1). A Praat script was used to extract the list of all labelled segments in each of the TextGrid objects for all corresponding sound files. These were saved as separate .csv files. Subsequently, all FACE tokens were visually inspected to determine the most appropriate number of formants required (4, 5 or 6 formants) for the Linear Predictive Coding algorithm to take plausible formant estimates. The selected number of formants required to take the most accurate formant measures for each vowel was then manually logged next to its corresponding label in the .csv files. Consistent spectrogram settings were applied throughout the analysis of all files, whereby the dynamic range was set to 45dB, in order to produce a spectrogram that displayed clear formants. The "maximum formant values" setting that resulted in the most appropriate measurements for each sound file (in this case, either 5000 or 5500 Hz) was also logged. Figure 4.2 displays the specific formant and spectrogram settings that were applied for the example token *maybe* (which can be seen in Figure 4.1).

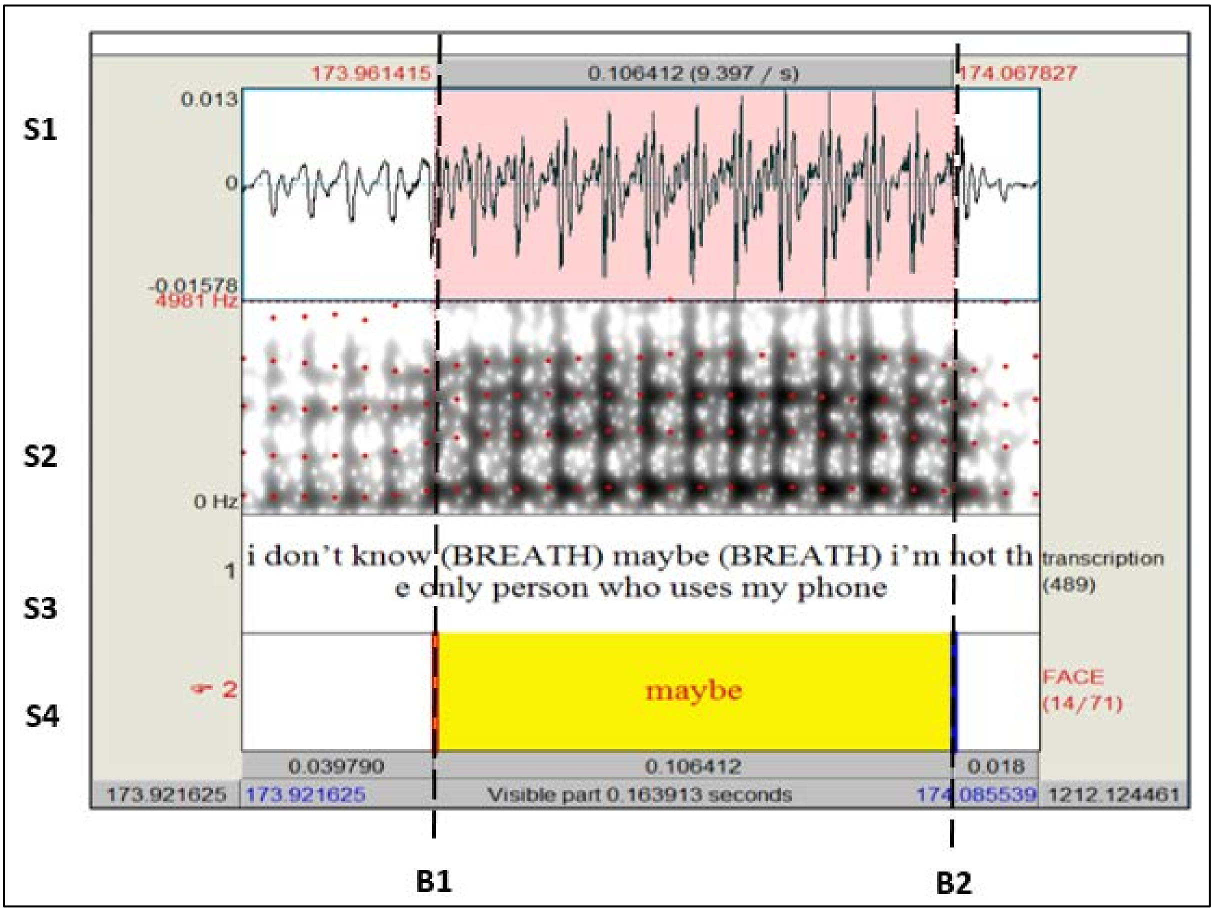


Figure 4.1. Praat window displaying extracted FACE vowel from the word *maybe.* Section 1 ('S1') shows the pressure waveform, 'S2' is the wide-band spectrogram, 'S3' is the 'transcription' tier containing all of the transcribed speech and 'S4' is the 'FACE' tier where the individual FACE vowel tokens were marked. Boundary 1 ('B1') marks the beginning of the FACE vowel and 'B2' marks the end of the vowel segment.

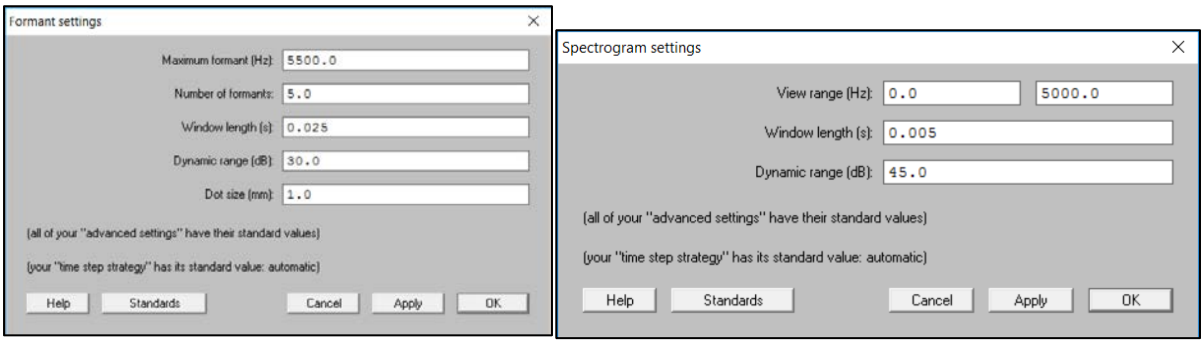


Figure 4.2. Praat windows displaying the formant settings (left) and spectrogram settings (right) that were applied for example token shown in Figure 4.1.

A Praat script (Lennes, 2017) designed to automatically extract measurements of the first three formants (F1~F3) at the midpoint of each labelled interval, was modified in order to measure the F1~F3 of the onset, midpoint and offset of the FACE vowel mid-section at 25%, 50% and 75% across the total vowel duration, respectively. The main advantage of using a proportional distance approach like this was that this method adjusts for the duration of each vowel and ensures that each vowel is handled consistently. Within the Praat script the formant settings could be specified according to the previously annotated requirements, including the number of formants and the maximum formant values. In a small number of instances (estimated to be around 1% of all tokens) where measurements based on the linear predictive coding algorithm appeared to be erroneous, manual corrections were made to ensure that the resonant frequencies were adequately represented based on visual inspection of the spectrogram. The formant data were subsequently exported to Microsoft Excel to be organised before statistical analysis was conducted in R (R Core Team, 2018).

4.4.3. Data processing

In total, formant values were measured and extracted from 2116 FACE tokens across all participants and tasks. The total number of tokens analysed across tasks and boroughs are summarised in Table 4.4. Due to the relatively short length of the Task 4 recordings, significantly less tokens were available in Task 4 than in Tasks 1 and 3. The average number of tokens selected from Tasks 1, 3 and 4 were 32, 29 and 9 respectively. The equivalent median values were 34, 31 and 9 respectively.

Table 4.4. Number of FACE tokens analysed, by borough and task.

	Number of FACE tokens analysed			
Area	**Task 1**	**Task 3**	**Task 4**	**All Tasks**
Bradford	316	301	92	709
Kirklees	311	262	89	662
Wakefield	340	318	87	745
All 3 boroughs	967	881	268	2116

FACE tokens were analysed from a range of phonetic environments and it was anticipated that this could have an influence on the acoustic properties under investigation. For this reason, care was taken to separate tokens into categories by phonetic context so that this factor could be accounted for when performing statistical analyses. In order to do this, all of the lexical items from which the FACE vowels had been extracted were checked and tokens from certain phonetic environments were assigned to specific groups. For example, tokens which occurred before or after a liquid were treated separately as liquids can have long-term resonance effects on adjacent formants and they can often cause lowering of F2 for front vowels (Ladefoged, 2001; Ladefoged & Maddieson, 1996). It was also recorded when tokens occurred next to either a glide or a nasal as these segments are acoustically similar to vowels as a result of being sonorant sounds produced with a relatively open vocal tract with a free air-flow (Rogers, 2000, p. 24). Rogers notes that although glides function as consonants, phonetically they are moving vowels (2000, p. 24). Additionally, FACE tokens from the words *make* and *take* (and *makes* and *takes*) were assigned to a separate group, as were tokens from words with <eigh> spellings. The justification for treating these sets of words separately was that previous literature has reported that the FACE vowel is often produced as the short variant [ɛ] in the words *make* and *take* in West Yorkshire and in words with <eigh> spellings, FACE usually becomes diphthongal [ɛɪ], even in areas where it would typically be a monophthong (Hughes et al., 2012, p. 105). The distribution of tokens across each phonetic environment is presented in Table 4.5, with each token analysed in this study appearing in only one group.

Table 4.5. Phonetic environments of FACE tokens.

Environment	Number of tokens	Assigned to group #
Between a liquid and a nasal	49	1
Preceding a nasal	278	2
Following a glide	133	3
Preceding a liquid	12	4
Following a liquid	473	5
From the words *make* and *take*	184	6
From words with <eigh> spelling	61	7
None of the above	926	0 *(default main group)*

Once all of the tokens had been labelled according to their phonetic environment and assigned to the appropriate group, a measure of how much tongue movement there was across each vowel token was taken. This was quantified by calculating the distances between the vowel onsets and offsets (measured at 25% and 75%, respectively) for F1 and F2. Greater distances indicated more tongue movement over the course of the vowel realisation and therefore more diphthongal articulations, while smaller distances indicated more stable, monophthongal realisations with less tongue movement during the vowel production. An increase in F1 would correspond to a vowel becoming more open while an increase in F2 would correspond to it becoming more fronted.

It should be noted that in sociophonetic research, it is often recommended that formant measurements of vowels should be normalised when comparing different speakers in order to determine the influence of demographic factors on variation (Di Paolo, Yaeger-Dror, & Beckford Wassink, 2011, p. 88). For sociophoneticians, the desired purpose of applying vowel normalisation techniques is to filter out physiological differences (such as differences in vocal tracts due to biological sex and aging) whilst leaving any sociophonetic differences intact (Di Paolo et al., 2011, p. 88). However, a number of scholars have taken the decision not to normalise their formant data in instances where the groups of speakers being compared are not judged to have considerably different vocal tracts, in terms of muscular settings and size (c.f. Devlin, 2014; Jacewicz, Fox, & Salmons, 2011). Furthermore, Devlin notes that while normalisation may be helpful when comparing speakers on a national or global scale, it is of less use in studies of local level regional variability where the speakers under investigation are not expected to demonstrate a contrast in vowel quality which would render vowel categories indistinguishable (2014, p. 99).

In the present case study, the formant data was not normalised as it was not deemed necessary to attempt to account for any physiological differences between participants because they were all of the same biological sex and were of similar ages, having all reached maturation. Additionally, it was anticipated that speakers from all three boroughs would have the same underlying phoneme for FACE. It is also worth noting that normalisation techniques are not generally applied in forensic phonetic research and as a primary aim of establishing the degree to which FACE varies across boroughs was to determine how narrowly or broadly reference populations need to be defined in forensic casework, this study adopted a similar approach. For these reasons, the raw midpoint formant values and distances between vowel onsets and offsets were used for all of the analyses reported in this chapter.

4.4.4. Statistical analysis

In addition to examining how the FACE vowel is realised across West Yorkshire, a key motivation of this study was to establish if FACE productions vary at a local level (i.e. across the boroughs of Bradford, Kirklees and Wakefield) and if so, to what extent. In order to test this, two separate sets of statistical analyses were carried out, each using different measurements relating to the FACE vowel. Firstly, the distance measures for F1 and F2 from the onset to the offset of the vowel were compared, in order to establish how much movement there was across the vowel and whether the amount of movement varied according to borough. Secondly, the values of the first three formants taken at the midpoint of each vowel token were compared across boroughs. Midpoint formant values were considered to be an adequate way of representing the vowel quality, as FACE is generally found to be monophthongal in West Yorkshire. However, it was also considered necessary to examine whether the degree to which the vowel was diphthongised varied according to borough, using the distance measures, prior to this analysis.

R version 3.5.0 (R Core Team, 2018) and lme4 version 1.1-21 (Bates, Mächler, Bolker, & Walker, 2015) were used to perform a series of linear mixed effects regression analyses. The first set of analyses tested the relationship between the movement of the vowel, as determined by the distance measures (offset formant values minus onset formant values), and the borough that the participants were from. It was necessary to perform a linear mixed effects analysis as this made it possible to add a mixture of fixed and random effects to the model. As a fixed effect BOROUGH was entered into the model, and this was treated as a categorical factor with three levels (Bradford, Kirklees and Wakefield, with Bradford as the reference level). The model also needed to account for any potential effects of varying speech styles across tasks and articulatory differences across phonetic environments from which FACE tokens had been collected. For this reason, TASK and PHONETIC ENVIRONMENT were also entered into the model as fixed effects, with TASK having three categorical levels (Task 1, Task 3 and

Task 4, with Task 4 as the reference level) and PHONETIC ENVIRONMENT having eight levels (one for each group listed in Table 4.5, with the largest category as the reference level). Task 4 was chosen as the intercept in the model because this was the baseline task and did not involve any interlocutors, whereas Bradford was treated as the intercept by default as it appears first alphabetically.

As multiple measures of the FACE vowel were taken from each participant, these responses could not be regarded as being independent from one another. In order to deal with this inter-dependency, a random effect for PARTICIPANT was entered into the model. This made it possible to resolve this non-independence by assuming a different "baseline" distance measure for each participant, whereby they were each assigned a different intercept value which had been estimated by the mixed model (Winter, 2013). It also needed to be taken into account that the effect of TASK may not be equal for all participants. Intuitively, it seems unlikely that all participants' FACE vowels would be influenced by the effect of TASK to the same degree because previous speech accommodation studies have found that accommodation behaviour can be conditioned by a wide range of factors relating to an individual. For instance, Yu, Abrego-Collier and Sonderegger (2013) reported that in their study, the extent of phonetic convergence in VOT was influenced not only by the listener's impression of the model talker, but also by listener traits including openness, conscientiousness and attention switching. For this reason, it was anticipated that while some participants may be relatively consistent across different speaking tasks, others might be more likely to converge towards or diverge away from their interlocutor and vary their speaking style to a strong degree across tasks. In order to take this into account, by-participant random slopes were included for the effect of TASK. The full linear mixed effects model with the F1 distance measures as the dependent variable is presented below:

F1.distance.model = lmer(DF1 ~ BOROUGH + TASK + ENVIRONMENT + (1+TASK|PARTICIPANT), data=Data, REML=FALSE)

In this model, the fixed effects of BOROUGH, TASK and PHONETIC ENVIRONMENT (without interaction terms) are used to predict F1 distance measures (i.e. F1 offset values minus F1 onset values). As random effects, there were intercepts for PARTICIPANT, as well as by-participant random slopes for the effect of TASK. This model was subsequently re-run using F2 distance measures as the dependent variable.

A further analysis was conducted to examine the relationship between the quality of the vowel (midpoint formant values) and the borough that the participants were from. For the same reasons as above, BOROUGH, TASK and PHONETIC ENVIRONMENT (without interaction terms) were entered into the model as fixed effects and as random effects, there were intercepts for PARTICIPANT as well as by-participant random slopes for the effect of TASK. The full linear mixed effects model with midpoint F1 values as the dependent variable is presented below:

F1.midpoint.model = lmer(F1 ~ BOROUGH + TASK + ENVIRONMENT + (1+TASK|PARTICIPANT), data=Data, REML=FALSE)

This model was subsequently re-run using F2 and then F3 midpoint values as the dependent variable. All models were fitted using maximum likelihood, and visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. In order to test the influence of each of the fixed effects on distance measures and midpoint formant values, respectively, p-values were obtained by likelihood ratio tests of the full model against the model without each of the fixed effects in question. Although the primary focus of this chapter is to examine regional variability in the FACE vowel, the influence of the task and phonetic environment are also reported and discussed in this chapter. The results of these likelihood ratio tests are presented in Section 4.5.2 below.

It should be noted that although other aspects of the FACE productions could have been analysed in this investigation, the static formant values were considered to be the best measure for characterising the FACE vowel quality in terms of being both comprehensive and straightforward. By collecting onset and offset measurements in addition to single-point midpoint formant values, it was possible to examine whether there were any differences in the amount of movement in FACE productions across the three boroughs. This is said to be the most commonly used multiple-measurement approach in sociophonetic research (Di Paolo et al., 2011, p. 91). As FACE was anticipated to be largely monophthongal in West Yorkshire, it was not considered necessary to map full dynamic formant trajectories. However, had this investigation involved a vowel phoneme that was expected to be realised as a diphthong by these speakers then more complex methods of vowel measurement and statistical analysis would have been required. For example, frequentist significance testing could have been conducted with dynamic formant trajectories using Generalised Additive Mixed Models (GAMMs; Sóskuthy, 2017; Wood, 2006).

4.5. Results

This section provides an overview of the distribution of FACE across West Yorkshire and presents the results of the linear mixed effects analyses in order to examine the extent to which any acoustic differences exist between FACE realisations across Bradford, Kirklees and Wakefield. The influence of the phonetic context and speaking task on FACE are also explored.

4.5.1. FACE in West Yorkshire

Although in this study the FACE tokens were not all individually transcribed phonetically, auditory impressions about the quality of the vowel were gathered whilst conducting the acoustic analysis. Using these auditory impressions and by taking into consideration the amount of movement from vowel onsets to vowel offsets, it was possible to establish how the FACE vowel is realised across West Yorkshire by these participants. Overall, it would seem that in West Yorkshire FACE is generally monophthongal and somewhere in the vicinity of [ɛ] or [e]. Impressionistically, FACE in Kirklees is generally close to [e̞] (average F1: 528 Hz, F2: 1704 Hz), Bradford is between [ɛ] and [e] (average F1: 558 Hz, F2: 1730 Hz), while Wakefield is closer to [ɛ̝] (average F1: 554 Hz, F2: 1803 Hz). However, it must be acknowledged that a range of variants for FACE including [eɪ], [ɛɪ], [æɪ], [eː], [ɛː] and [ɪ] were all observed within each of the three boroughs.

For each of the three boroughs considered in this investigation, the average midpoint F1~F3 values for FACE were calculated as well as their standard deviations. The average distance measures from vowel onset to offset (measured at 25% and 75%, respectively) were also measured for F1 and F2 across all participants from each of the three boroughs in addition to the standard deviations of these values. Table 4.6 provides a summary of the distribution of both the midpoint formant values and the distance measures across the boroughs of Bradford, Kirklees and Wakefield, as well as the three areas together. The acoustic analysis of the formant data shows that there is relatively little movement from vowel onset to vowel offset, particularly in terms of F1. More movement is present across the vowel in the second formant; however, the majority of FACE vowel tokens were generally perceived to be realised as monophthongs auditorily.

Table 4.6. Summary of FACE formant values and distance measures from vowel onset to offset, across the boroughs of Bradford, Kirklees, Wakefield and all three areas together.

		Midpoint formant values (Hz)			**Distance measures (Hz)**	
Area	**Measure**	**F1**	**F2**	**F3**	**F1**	**F2**
Bradford (N=709)	**Average**	558	1730	2483	-16	105
	SD	55	144	158	64	153
Kirklees (N=662)	**Average**	528	1704	2500	-16	97
	SD	45	158	181	51	133
Wakefield (N=745)	**Average**	554	1803	2561	-40	141
	SD	59	171	166	63	152
All 3 Boroughs (N=2116)	**Average**	547	1747	2516	-24	115
	SD	55	164	172	61	148

Figures 4.3 presents the average trajectories from vowel onset to offset, for all 30 participants. Each participant's trajectory is colour coded according to the borough that they are from. In this figure, slight separation can be seen across boroughs both in terms of F1 and F2; however, the regional differences on the F2 dimension appear to be strongest, indicating that vowel front/backness is perhaps most regionally marked. It can be seen that, overall the trajectories for the Kirklees participants are more close and further back than those of the Bradford and Wakefield participants. For the majority of participants, the trajectories are

relatively small. In terms of the direction of the trajectory, F2 consistently increases across the vowel (meaning that the vowel becomes more front) and, for the majority of participants, F1 decreases across the vowel (meaning that the vowel becomes more close over the duration of the articulation). There were six participants for whom their average F1 was higher at the vowel offset than the vowel onset, although in all of these cases the average distance measure was less than 15 Hz.

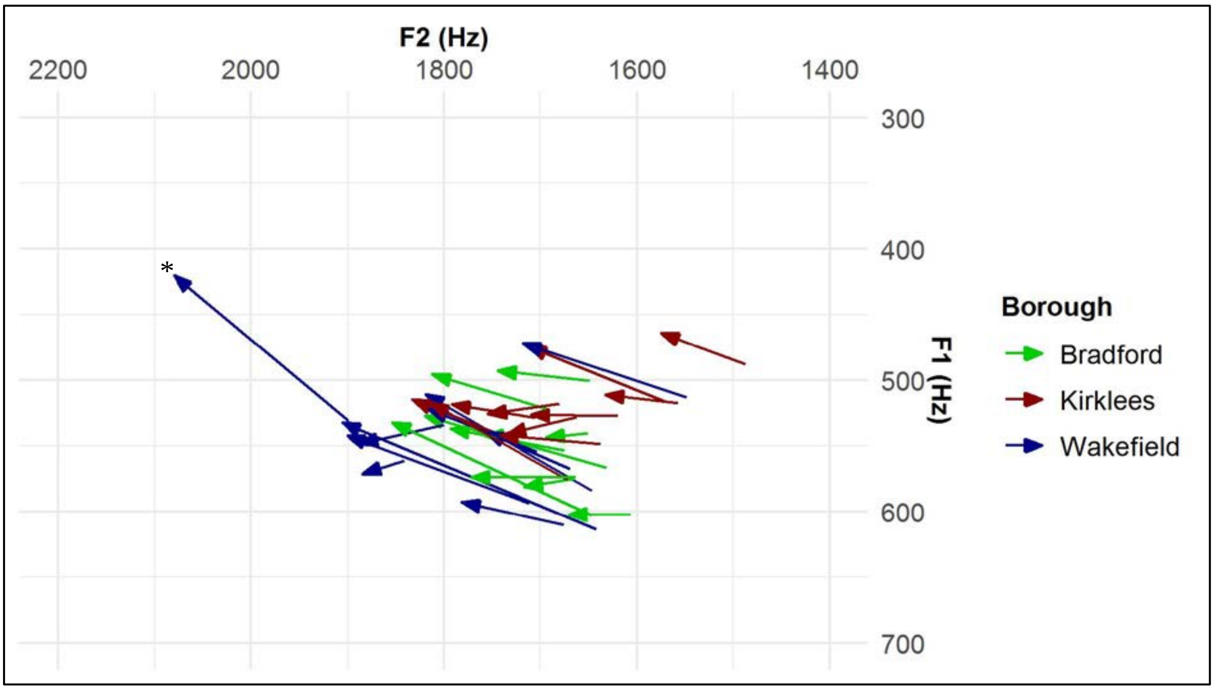


Figure 4.3. Average F1 and F2 values at 25% and 75% of the FACE vowels plotted for all 30 participants.

In Figure 4.3, it is possible to see quite clearly that one of the Wakefield participants (participant #041, indicated with *) has a much more front and close average vowel offset than the rest of the participants. This participant also appears to have one of the largest average FACE vowel trajectories, indicating that his FACE tokens may be more diphthongal than some of the other participants, in Wakefield, and West Yorkshire more broadly. In order to find an explanation for this, participant #041's FACE tokens in all three files were analysed auditorily and the formant readings were double checked. As studies have shown that there is a systematic relationship between f_0 and formant frequencies, with average f_0 and average F1~F3 being moderately correlated (Assmann, 2008), this participant's f_0 was also taken into account. It was observed that this participant had a relatively high average f_0 of approximately 140 Hz which may partly explain why the F2 values were so high. Furthermore, he also tended to produce particularly diphthongal realisations of FACE in certain phonetic environments, such as when FACE occurred after a glide in tokens such as *way* (Task 1, 311s), *Rachel* (Task 1, 571s) and *brakes* (Task 1, 1870s). However, the majority of FACE tokens produced by this participant were classified as monophthongal and were not dissimilar to the other Wakefield participants' realisations. As the trajectories plotted in Figure 4.3 represent an average onset and offset based on all of the tokens that each participant has produced, tokens following a glide may have skewed the results. Additionally, the average midpoint formant values and distance measures presented in Table 4.6 for the Wakefield borough may be slightly skewed as a result of this participant. However, it should be noted that the linear mixed effects models accounted for the fixed effects of PHONETIC ENVIRONMENT and included random effects for PARTICIPANTS, and therefore this participant was assigned an appropriate baseline intercept, as required to take this into account.

The boxplots in Figure 4.4 visualise the midpoint formant data for all three boroughs. It can be seen that the distribution of FACE formant values varies across Bradford, Kirklees and Wakefield. Again, it must be acknowledged that some of the variation may be caused by slightly unbalanced distributions of FACE tokens from different phonetic environments and there being some occasional diphthongal tokens. However, it was not possible to separate monophthongs from diphthongs in this study, as this was primarily an acoustic study and there is no clear-cut boundary between the two. Please note that the y-axes for each of the F1~F3 subplots have been customised within a range the best suits the formant values and therefore they are all different. Full details of the F1~F3 lower quartiles, medians, upper quartiles and interquartile ranges are provided in Table 4.7.

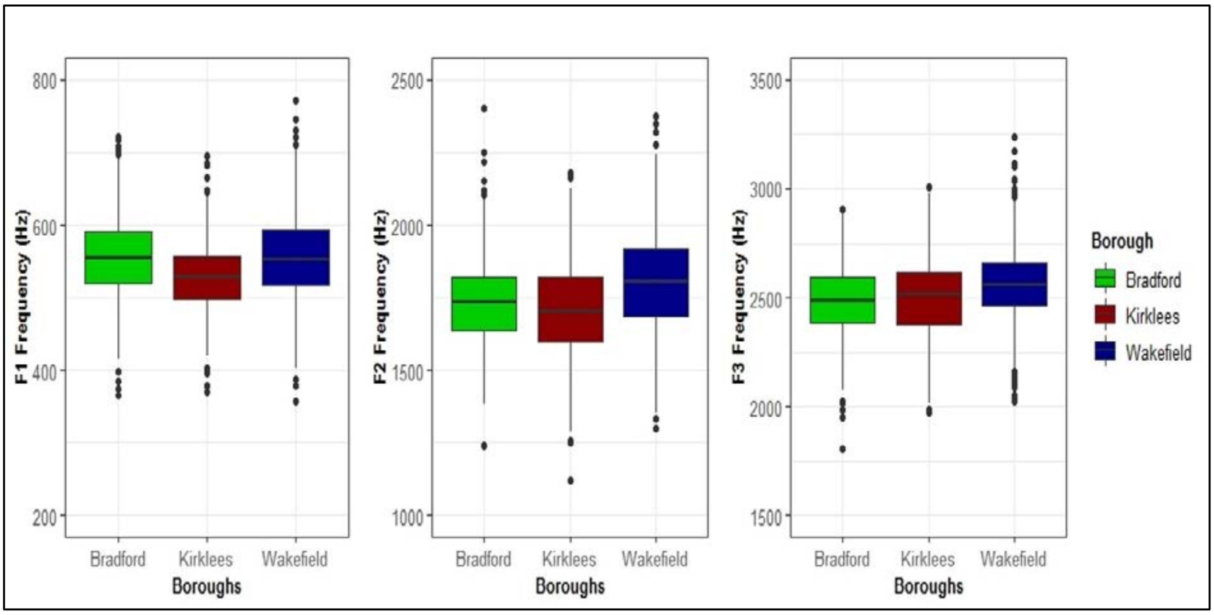


Figure 4.4. Midpoint F1~F3 data across West Yorkshire boroughs.

Table 4.7. Lower quartile, median, upper quartile and interquartile range for F1~F3 midpoint formant values across the boroughs of Bradford, Kirklees and Wakefield.

		Midpoint formant values (Hz)		
Measure	**Borough**	**F1**	**F2**	**F3**
Lower Quartile	Bradford	520	1636	2385
	Kirklees	497	1600	2376
	Wakefield	517	1685	2464
Median	Bradford	555	1731	2488
	Kirklees	527	1703	2513
	Wakefield	552	1805	2558
Upper Quartile	Bradford	591	1821	2592
	Kirklees	556	1822	2620
	Wakefield	593	1918	2659
Interquartile Range	Bradford	71	184	207
	Kirklees	59	222	243
	Wakefield	76	233	195

Figures 4.5-4.7 present kernel density plots for all of the midpoint F1~F3 values, respectively. These plots visualise the distribution of the FACE vowel formant data, by overlaying colour-coded kernel smoothed plots for each of the three West Yorkshire boroughs. Please note that the x-axes for each of density plots have been customised within a range the best suits the formant values and therefore they are all different. Based on the results presented in this section, it would appear that F2 varies the most across the three boroughs. F1 and F3 both seem to vary but to a lesser extent. However, it must be recognised that these figures take into account all tokens from across all participants, tasks, and phonetic environments. In order to ensure that comparisons across boroughs are comparing like with like, the results of the linear mixed effects analyses need to be taken into account.

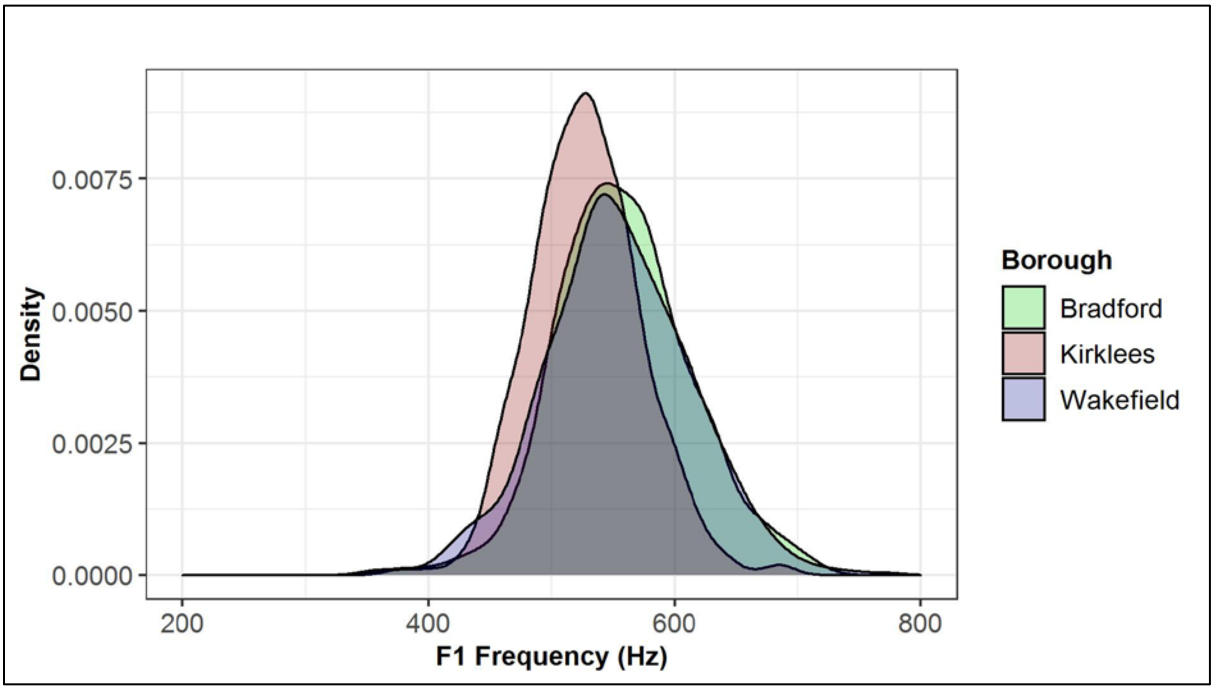


Figure 4.5. Distribution of F1 midpoint formant data across West Yorkshire boroughs.

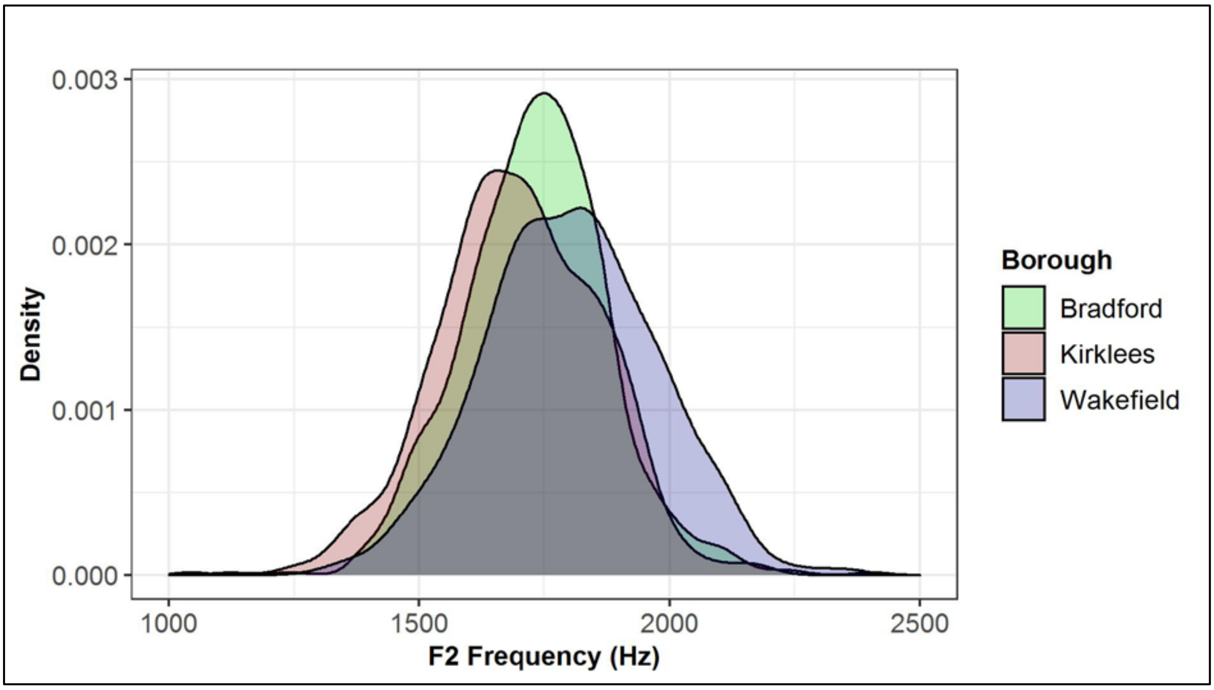


Figure 4.6. Distribution of F2 midpoint formant data across West Yorkshire boroughs.

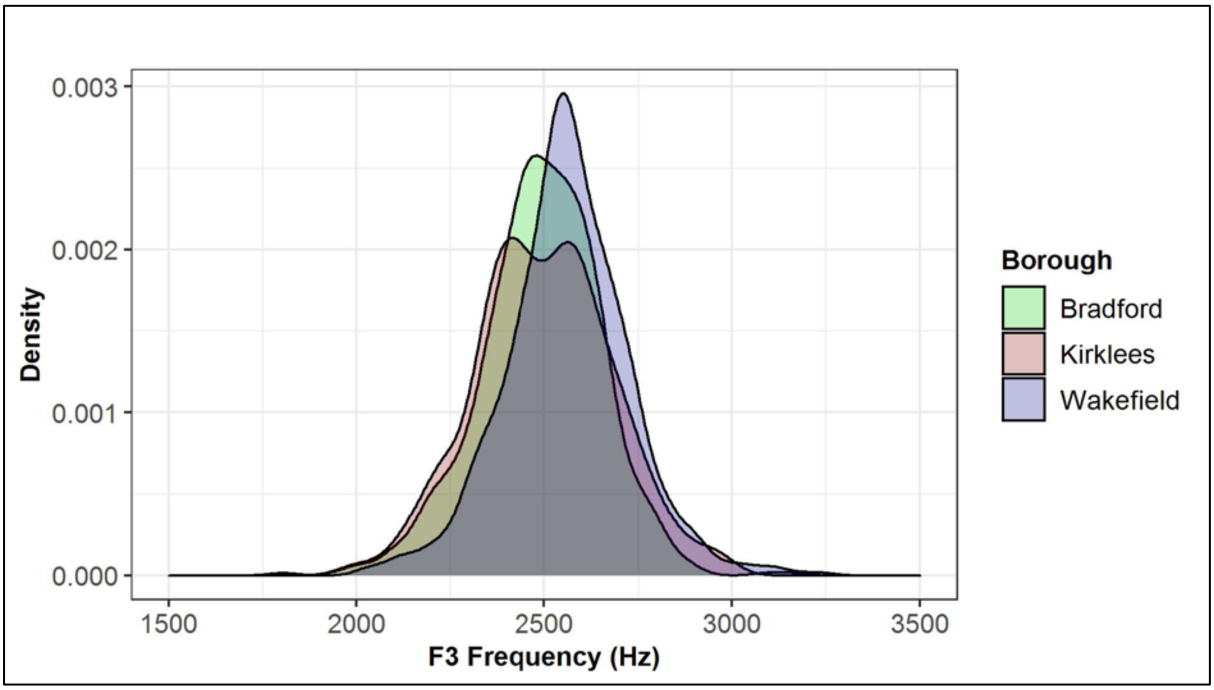


Figure 4.7. Distribution of F3 midpoint formant data across West Yorkshire boroughs.

4.5.2. Linear mixed effects analyses results

4.5.2.1 Vowel movement

4.5.2.1.1 Influence of BOROUGH

The first set of linear mixed effects analyses described in Section 4.4.4 compared the distance measures for F1 and F2 across the boroughs of Bradford, Kirklees and Wakefield. The results of these analyses showed that there were no significant differences across boroughs for F1 or F2 distance measures ($p > 0.05$ in both cases). This meant that, when taking into account all tokens rather than comparing averages (as in Figure 4.3), the movement in the FACE trajectories from vowel onset to offset did not significantly vary as a result of which borough the participants were from. This was possibly due to the fact that the FACE vowel was generally monophthongal throughout the three West Yorkshire boroughs. As a consequence of this finding, it was deemed appropriate to focus on the midpoint F1~F3 formant values when assessing how the quality of the FACE vowel varies between boroughs.

4.5.2.1.2 Influence of TASK and PHONETIC ENVIRONMENT

With respect to the control variables TASK and PHONETIC ENVIRONMENT, the likelihood ratio test results showed that the distance measures did not vary significantly across the three tasks; however, PHONETIC ENVIRONMENT was a significant predictor in both of the models. PHONETIC ENVIRONMENT significantly affected F1 distance measures ($\chi^2(7) = 309.38$, $p < 0.001$) and F2 distance measures ($\chi^2(7) = 880.55$, $p < 0.001$) with high levels of variability present across the eight categories. These results suggest that variations in the amount of movement associated with FACE productions can be better explained by the phonetic context of the FACE token, rather than by the effect of the task or where the participant is from. It should be noted that the F1 distance measure model failed to converge when including by-participant random slopes for the effect of TASK, indicating that the model was too complex for the available data. For this reason, the model was re-run without the by-participant random slopes and this time the model reached convergence. The findings of the likelihood ratio tests remained the same, with PHONETIC ENVIRONMENT being the only fixed effect to have a significant influence on F1 distance measures. Full summary tables for the F1 and F2 distance models are included in Appendix 3.

Figures 4.8 and 4.9 illustrate the variation in distance measures across phonetic environments for F1 and F2, respectively. These violin plots are similar to boxplots, except that they also show the kernel probability density of the data at different values. Please note that the y-axes for each of these figures has been customised within a range the best suits the formant distance measures and therefore they are not all the same. In these figures, the further the distance measure from zero the more movement there was from onset to offset of the vowel.

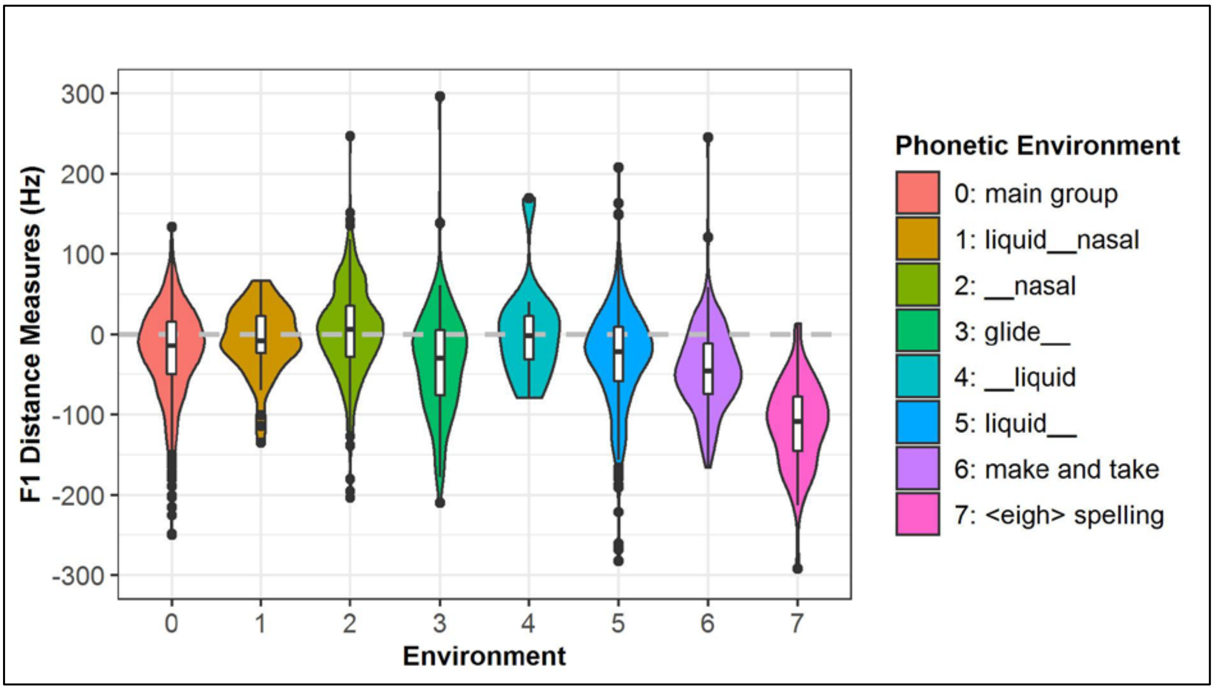


Figure 4.8. Distribution of F1 distance measures across phonetic environments.

In Figure 4.8, it can be seen that there is quite a large amount of variation across phonetic environments in terms of the shape and spread of each distribution, however, the median values are largely close to 0. The main exceptions to this are in groups 6 and 7, corresponding to FACE tokens occurring in the words *make* and *take* and in words with <eigh> spelling, in which F1 typically decreased over the course of the tokens. This increase is particularly large in words with <eigh> spelling, suggesting a typical realisation close to [ɛɪ]. It is perhaps unsurprising that the distance measures were generally furthest from 0 in FACE tokens from words with <eigh> spellings, as diphthongal variants are often found in this context (Hughes et al., 2012; Rogers, 2000; Stoddart et al., 1999). The F1 distance measures in FACE tokens from the default main group (group 0) mainly range between reductions of 100 Hz to increases of

100 Hz over the course of the tokens, with a median F1 distance measure close to 0. However, there are also examples of individual tokens in this group in which the F1 distance measures have decreased by as much as 250 Hz, representing quite substantial raising of the tongue during these FACE productions. The F1 distance measures from FACE tokens occurring after a liquid (group 5) also reflect a large amount of movement in the tongue height dimension, ranging from a reduction in F1 of approximately 250 Hz up to an increase in F1 of 200 Hz.

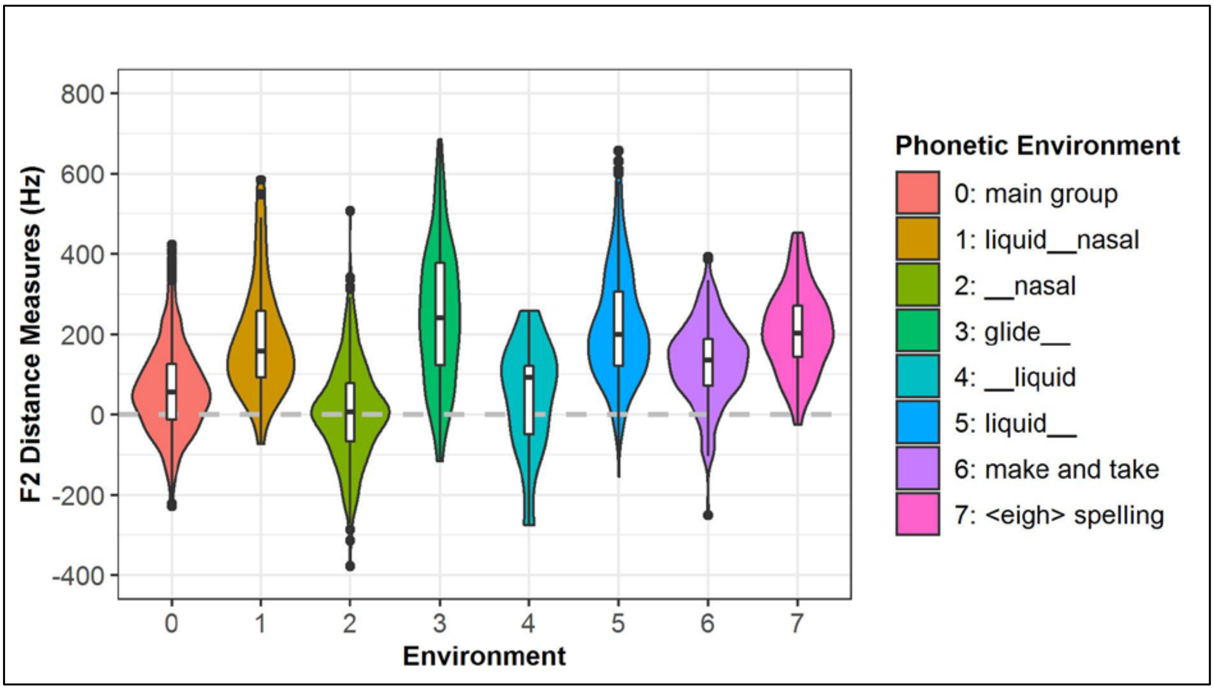


Figure 4.9. Distribution of F2 distance measures across phonetic environments.

In Figure 4.9, a general pattern is evident whereby FACE F2 values tend to increase from vowel onset to offset across all phonetic contexts. This corresponds to the tongue moving further forward in the mouth over the course of the FACE vowel production. However, the extent to which F2 increases is highly variable across the different contexts. Compared to the default main group, FACE tokens occurring after a glide or a liquid or in words with <eigh> spellings (groups 3, 5 and 7) stand out as having the most marked increase in F2 distance measures. In contrast, tokens produced before a nasal (group 2) have much smaller F2 distance measures overall, whereas the FACE tokens produced before a liquid (group 4) have similar F2 distance measures to the main group. Overall, the distance measures visualised in Figures 4.8 and 4.9 suggest that there was more movement from FACE vowel onset to offset in terms of F2 than

F1, and there was also more variation across phonetic environments in F2 distance measures than in F1 distance measures.

4.5.2.2 Vowel quality

4.5.2.2.1 Influence of BOROUGH

The second set of linear mixed effects analyses compared the midpoint F1~F3 values across the boroughs of Bradford, Kirklees and Wakefield, taking into account by-participant variation. The results of these analyses revealed that the vowel quality of the participants' FACE productions varied at a local level across the three boroughs. Taking Bradford as the intercept in the models, the results of the linear mixed effects analyses showed that BOROUGH significantly affected F2 ($\chi^2(2)=10.34$, $p<0.01$), lowering it by about 26 Hz ± 33 (standard errors) for Kirklees and increasing it by about 90 Hz ± 33 (standard errors) for Wakefield. F1 and F3 values did not vary significantly across the three boroughs. Again, it should be noted that the F1 model failed to converge and subsequently the F1 model was re-run without the by-participant random slopes and this time the model reached convergence. The effects of BOROUGH remained the same in the updated model.

4.5.2.2.2 Influence of TASK

In relation to the effects of TASK, the likelihood ratio test results showed that there were significant differences in terms of F2 values across the three tasks overall. Taking Task 4 as the intercept in the models, TASK significantly affected F2 ($\chi2(2) = 15.04$, $p < 0.001$), lowering it by about 57 Hz ± 14 (standard errors) for Task 1 and lowering it by about 60 Hz ± 13 (standard errors) for Task 3. The likelihood ratio test results showed that F1 and F3 values did not vary significantly across the three tasks. When the models were re-run without the by-participant random slopes, the findings of the likelihood ratio tests comparing the updated F1 model with and without the fixed effect of TASK showed that TASK significantly affected F1 ($\chi^2(2) = 14.26$, $p < 0.001$), increasing it by about 10 Hz ± 3 (standard errors) for Task 1 and increasing it by about 3 Hz ± 3 (standard errors) for Task 3. Despite the change in results when using the revised F1 model, it was felt necessary to account for the fact that the effect of TASK may not be equal for all participants and therefore only the differences in F2 values were considered to be significantly different overall.

4.5.2.2.3 Influence of PHONETIC ENVIRONMENT

PHONETIC ENVIRONMENT significantly affected F1 ($\chi^2(7) = 79.77$, $p < 0.001$), F2 ($\chi^2(7) = 713.9$, $p < 0.001$), and F3 ($\chi^2(7) = 472.45$, $p < 0.001$), with high levels of variability present across the eight categories. This was also the case when the model was re-run without the by-participant random slopes. Full summary tables for the midpoint F1~F3 models are included in Appendix 3. Figures 4.10-4.12 illustrate the variation in midpoint F1~F3 values across phonetic environments, respectively. Again, the y-axes for each of these plots has been customised within a range the best suits the formant values and therefore they are not all the same.

In Figure 4.10, it can be seen that the median F1 midpoint values of FACE were largely around 550 Hz across all of the phonetic environments considered in this study. Overall, the majority of F1 values for the default main group (group 0) range from approximately 400 Hz up to 700 Hz. This was also the case for FACE tokens that occurred before a nasal, after a glide and after a liquid (groups 2, 3 and 5). Compared to the main group, FACE tokens in <eigh> words tended to have lower F1 values whereas tokens occurring before a liquid tended to have higher F1 values. With reference to Figure 4.11, it would appear that there were higher levels of variability in F2 values across phonetic environments than in F1 values. The majority of F2 values for the main group range from approximately 1500 Hz up to 2200 Hz. The greatest deviations from the F2 distributions of the main group were in FACE tokens following glides and liquids (groups 3 and 5) where F2 values were generally lower, and in <eigh> words (group 7) where F2 values were much higher.

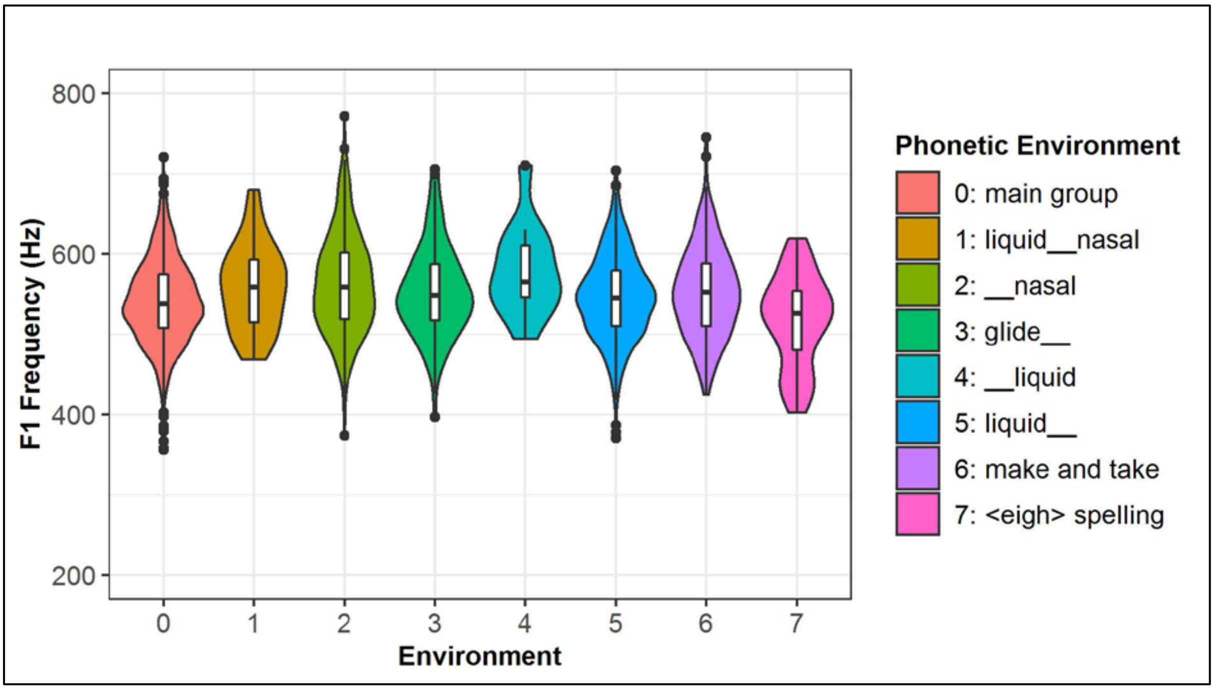


Figure 4.10. Distribution of F1 midpoint formant data across phonetic environments.

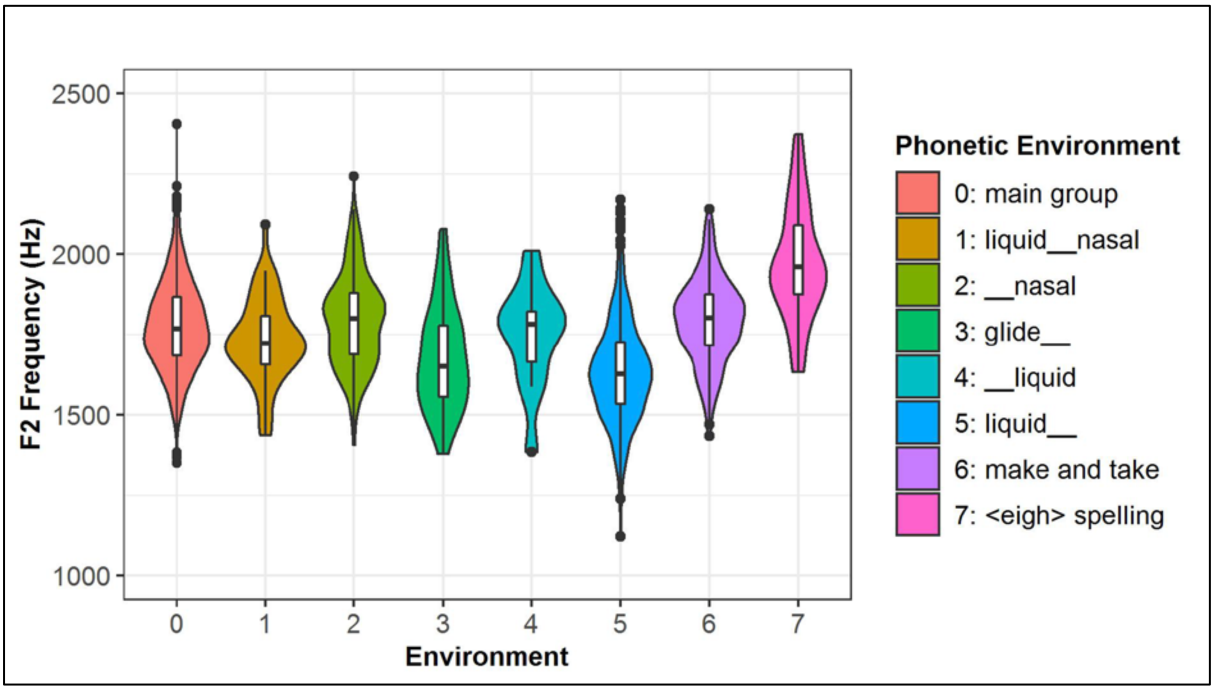


Figure 4.11. Distribution of F2 midpoint formant data across phonetic environments.

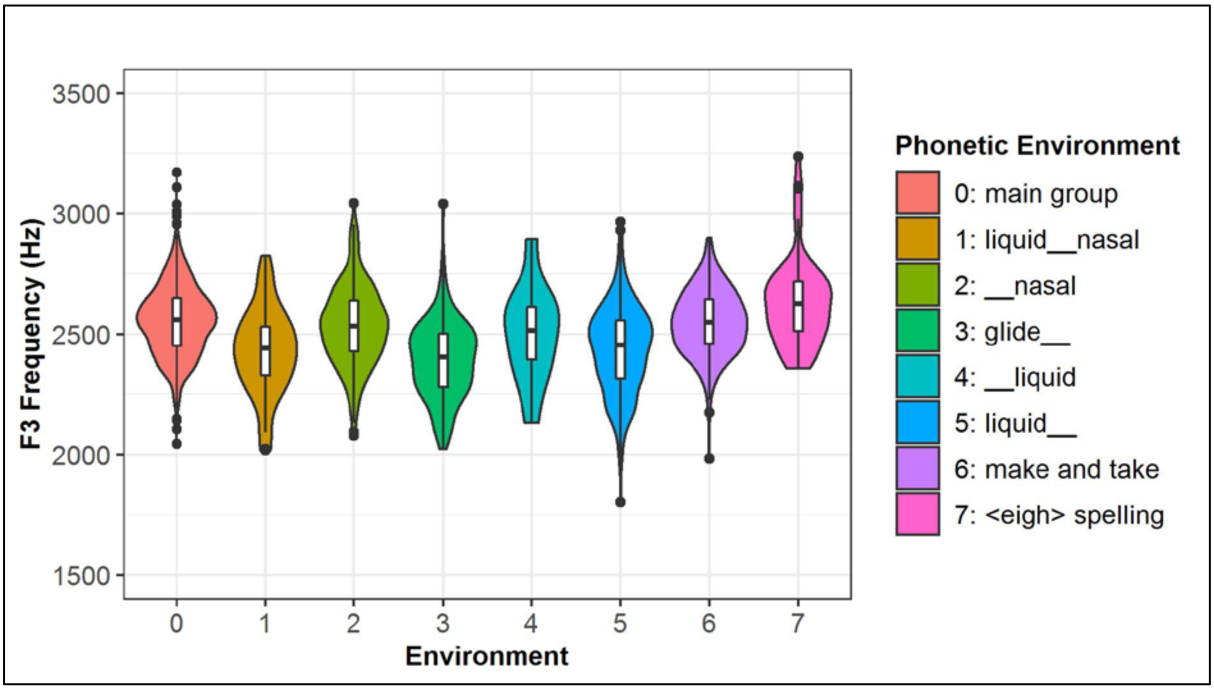


Figure 4.12. Distribution of F3 midpoint formant data across phonetic environments.

Figure 4.12 illustrates that there was also some variation in F3 across the different phonetic environments in which FACE was produced. Based on median values, F3 is generally higher in the main group than in all of the other phonetic environments except for the FACE tokens in group 7. In the main group, F3 values mostly ranged from 2250 Hz to 3000 Hz, whereas FACE tokens occurring in <eigh> words had F3 values as high as 3250 Hz. Overall, the findings in relation to how the midpoint formants of FACE varied across the different phonetic environments examined in this study highlight the need to consider the surrounding phonetic context when analysing variation in vowel productions.

In summary, the results revealed that the FACE vowel front/backness varied according to which borough the participants were from. In contrast, F1 and F3 midpoints were not significantly affected by the fixed effect of BOROUGH. Results also indicated that FACE realisations are influenced by the phonetic context in which the token is produced as well as the effect of speaking task. In relation to the effects of TASK, the F2 dimension of FACE relating to vowel front/backness may be most susceptible to the effects of speech style and speech accommodation, however, this will be explored in more depth in the following chapter.

4.6. Discussion

This section begins by recapping previous findings in relation to FACE in West Yorkshire and then goes on to compare them with the findings of the present study. The results of this investigation are then discussed in more detail and subsequently the implications of the findings for researchers in the fields of forensic speech science and sociophonetics are highlighted.

4.6.1. Summary of findings

The existing literature on West Yorkshire English had previously indicated that the production of FACE varied across the region. Descriptions of auditory impressions of the FACE vowel in Bradford and Kirklees included [ɛɪ] and [eː] as common variants (Petyt, 1985, p. 120). Bradford English was also said to typically contain an open-mid monophthong [ɛː] (Hughes et al., 2012, p. 105). Whereas in Wakefield, FACE was reported to be most commonly realised as [eː] (Burland-Gibson, 2019, p. 190). The present study illustrates that it is still the case that regional variation exists across the boroughs of West Yorkshire with respect to FACE, by using evidence obtained from both auditory and acoustic analyses.

In this study, auditory inspections of the FACE vowel in the boroughs of Bradford, Kirklees and Wakefield showed that generally speaking FACE was overwhelmingly monophthongal and somewhere in the vicinity of [ɛ] or [e]. Impressionistically the precise vowel realisations varied slightly according to borough; however, it was acknowledged that a range of variants including [ɛ], [ɛː], [eː], [ɪ], [eɪ], [ɛɪ] and [æɪ] were observed within each of the three boroughs, when considering all tokens from all participants. In comparison to previous descriptions of the FACE vowel in the literature, it would seem that a wider range of FACE variants are in use across the boroughs of Bradford, Kirklees and Wakefield than were previously cited. However, the majority of the studies referred to in this chapter did not provide specific details of the phonetic context in which the FACE vowels were selected and therefore it is unclear how comparable the methodological designs were. Descriptions of FACE were mostly provided as part of a general summary of the phonetic features of a particular accent, rather than an in-depth analysis of that specific vowel. Therefore, it could be the case that a narrow range of FACE variants were observed in previous studies because only a subset of the phonetic environments included in this investigation were analysed. Furthermore, as no acoustic analyses were undertaken in the earlier studies of FACE in West Yorkshire, it is not possible to comment on how FACE may have changed in terms of specific formant values. Nevertheless, it would appear that generally speaking, FACE in West Yorkshire is still largely monophthongal and regional variation is still present at the local borough level.

Based solely on auditory perceptions, it would be difficult to assess the exact degree to which the FACE vowel varies across the three West Yorkshire boroughs, as the differences in realisation between boroughs are fairly subtle. However, when taking into account the acoustic information relating to the first three formants of the FACE vowel, it became clear that regional variation exists and it was possible to substantiate these findings using statistical methods. The linear mixed effects models and likelihood ratio test results illustrated that F2 is more regionally influenced than F1 and F3. It was determined that significant differences existed between boroughs in terms of vowel front/backness, with F2 being highest in Wakefield and lowest in Kirklees. These trends were similarly reflected by the average and median values for the F2 midpoints across the three boroughs, presented in Tables 4.6 and 4.7, respectively.

In terms of F1 and F3, the linear mixed effects analyses revealed that there were no significant differences across the boroughs of Bradford, Kirklees and Wakefield. One explanation for why F3 may vary the least across boroughs is that this parameter is generally considered to be more idiosyncratic (Gold, French, & Harrison, 2013b; Hughes, McDougall, & Foulkes, 2009; McDougall, 2004) and therefore less likely to be dependent on external factors, such as the place the speaker is from. Differences in F3 may also be harder to perceive and less consciously controlled by the speakers. When considering the average values for F3 midpoints across the three boroughs, it can be seen that there was less than 100 Hz difference between Bradford (the borough with the lowest average F3) and Wakefield (the borough with the highest average F3). The standard deviations of the F3 midpoint values were also relatively similar across the three boroughs (158, 166 and 181 Hz in Bradford, Wakefield and Kirklees respectively).

A further finding of this study was that the degree to which the FACE vowel was diphthongised did not appear to be regionally stratified. It was observed that the distance measures from vowel onset to offset for F1 and F2 were not good predictors of which borough the participants were from (i.e. there were no significant differences in these values across the three boroughs). The reason for this may be due to the fact that the vast majority of FACE tokens are classified as monophthongal and any tokens that were diphthongal displayed a relatively small amount of movement across the vowel, mainly in terms of F2. The average F1 and F2 distance measures for the three boroughs, presented in Table 4.6, show that participants from Wakefield had the most movement from vowel onset to vowel offset overall. Although, as was previously mentioned in Section 4.5.1, participant #041 is likely to have skewed the average Wakefield values as he had a particularly large average vowel trajectory. Participants from Kirklees seemed to have the most monophthongal FACE vowels with the shortest distance from vowel onset to offset for F1 and F2. However, when considering the precise values of the distance measures, it is clear to see that FACE was generally monophthongal with relatively little movement across the vowel in all three boroughs. This was also evident when looking at the average vowel trajectories shown in Figure 4.3.

4.6.2. Implications

4.6.2.1 Forensic speech science

A vital part of FSC casework involves making an assessment of how typical a particular speech parameter is in a given population. Experts make such typicality judgements based on their knowledge and experience of speakers from the relevant population, and also with reference to existing population data and academic research findings. Although this is a relatively small-scale study, in that it only considers 30 speakers, it is intended that this may serve as a useful resource for forensic caseworkers which can add to the bigger picture of local level variability in West Yorkshire. Specifically, this study informs forensic caseworkers of how the FACE vowel is realised across West Yorkshire by providing descriptions of FACE variants found in the boroughs of Bradford, Kirklees and Wakefield and it also provides new vowel formant data for an area for which limited acoustic data was previously available. By referring to the results of this investigation, caseworkers can gain a better idea of what to expect in this region in terms of the FACE vowel and can use this information when conducting FSC casework involving West Yorkshire speakers. For example, if a speaker who was believed to be from West Yorkshire used a wide diphthongal variant for FACE, such as [æi], the expert would be able to compare this to the findings of this study and make an informed judgement that this was an unusual FACE variant, for this particular speech community. Equally, if they were to observe a monophthongal FACE variant in the region of [ɛ] or [e], they would be able to say that this was relatively typical for West Yorkshire English.

The vowel formant data presented in this chapter could also be used to make an assessment of how typical a particular set of FACE tokens were in a forensic sample. For instance, using the density plots shown in Figures 4.5-4.7, it is possible to estimate the range in which F1~F3 values typically fall for FACE. In Figure 4.5 it can be seen that FACE F1 midpoint values are most heavily concentrated within the region of 500 to 600 Hz and Figure 4.6 shows F2 values largely tend to be between 1600 and 1900 Hz. Therefore, if a young male from West Yorkshire being analysed in a FSC case displayed FACE tokens with F1 values of around 550 Hz and F2 values close to 1750 Hz, this would be considered to be a relatively typical FACE production. Whereas a West Yorkshire male displaying F1 values in the region of 750 Hz and F2 values around 1200 Hz might be considered to have unusual realisations of FACE for this population. In a FSC case, if both the suspect and offender data fell near the less typical, latter frequencies this would provide evidence in support of the same speaker view; that the evidence in the speech samples would be more likely to occur if it had been produced by the same person. However, further data would be required in order to adequately estimate the strength of the evidence.

Although the vowel formant data is useful for forensic phoneticians, it must be noted that caution is required when comparing raw formant values, as physiological factors can affect vowel formant frequencies and therefore these must be taken into account. For example, if we were to compare the F1~F3 values reported in this chapter with those of a female speaker from West Yorkshire's FACE vowel, we would expect the female speaker to have much higher formant values as a result of differences in vocal tract length compared to the male participants in this study. Furthermore, it may be the case that females generally pattern differently for the area than males. It must also be acknowledged that other factors such as age, ethnicity and social class have been shown to influence a person's accent, and therefore it is unlikely that these results are fully representative of West Yorkshire. It must be noted that all of the participants in this study are white, university educated, males aged 19-29. Ideally, it would be useful to repeat this study using groups of speakers of different ages, ethnicities and social classes. However, time and financial restrictions make this prohibitively difficult and this is why there is such a scarcity of reference population data available for forensic phoneticians to consult when carrying out casework. While it is possible that the findings presented in this chapter may be applicable to speakers from a range of other sociodemographic groups, the degree to which these results can be extrapolated across the whole of West Yorkshire is unknown.

By considering the extent to which FACE varies across the areas of Bradford, Kirklees and Wakefield it is possible to draw conclusions about how narrowly reference populations need to be delimited. The findings of this study highlight that there is regional variation present on a more local level than one might expect as the results of the linear mixed effects analyses suggest that there are significant acoustic differences in FACE F2 frequencies between the boroughs of Bradford, Kirklees and Wakefield. Furthermore, auditory perceptions of the FACE vowel confirmed that the precise vowel realisations appeared to vary slightly according to borough. This evidence indicates that it might be more appropriate to treat the West Yorkshire boroughs as separate populations rather than being grouped together more generally as West Yorkshire speakers. When taking the density plots presented in Figures 4.5-4.7 into account, it can be seen that there is a substantial amount of overlap between the FACE vowel distributions across the three West Yorkshire boroughs, however the peaks appear distinct.

With regards to F1, Figure 4.5 shows that all three boroughs display roughly normal distributions and it can be seen that the Bradford and Wakefield formant values are very similarly distributed. Kirklees F1 values are slightly lower and appear to fall within a narrower range than those of Bradford and Wakefield and therefore it could be argued that this borough should be treated separately with respect to F1 values. Similarly to F1 values, F3 values seem to overlap quite heavily across Bradford, Kirklees and Wakefield. Figure 4.7 shows that F3 is normally distributed for both Bradford and Wakefield but there are two peaks in the Kirklees density plot, meaning that this distribution is bi-modal. The majority of

Wakefield values fall within a narrower range than those of Bradford and Kirklees but overall the F3 values are most heavily concentrated in the region of 2250 to 2750 Hz. In terms of F3 values it would seem reasonable to group the three separate boroughs together as one West Yorkshire group. Although there is still significant overlap, F2 distributions show the most separation between the three boroughs. All boroughs show a broadly normal distribution; however, Kirklees is slightly positively skewed whereas Wakefield is slightly negatively skewed. Based on the distribution of F2 values displayed in Figure 4.6, it would seem that the three boroughs should be treated as separate speech communities.

The results of this study indicate that, for FACE at least, attention must be paid to local level variation as significant acoustic differences exist across West Yorkshire even within a fairly homogeneous community. Based on the 30 participants in this investigation, it would seem reasonable to treat the boroughs of Bradford, Kirklees and Wakefield as separate populations when assessing the strength of evidence in order to avoid over- or under-estimations. This is useful for forensic phoneticians to be aware of when delimiting the relevant reference population and when making assessments of the strength of the evidence in FSC casework.

4.6.2.2 Sociophonetics

In addition to the implications that this investigation may have on forensic speech science research and casework, this study also provides insights into an area which has seldom been analysed from a sociophonetic perspective. The findings of this investigation are important for sociophoneticians as they provide an up-to-date account of FACE in West Yorkshire, which includes auditory and acoustic data. Additionally, this study demonstrates the benefits of using both auditory and acoustic data when examining variation across different speaker groups. In the case of the FACE vowel, acoustic information in the form of formant data made it possible to identify and quantify fine-grained articulatory differences across boroughs, which were not as easily observable using auditory analysis only. For this reason, it is considered advisable to use both auditory and acoustic methods when conducting sociophonetic research, where possible.

One explanation for the regional variation in the FACE vowel that has been observed in this study, could be that this phonetic parameter is being used to signal different versions of a local West Yorkshire identity. The fact that the use of [tɛk] for *take* is being employed in local advertisements to index a stereotypical West Yorkshire identity (cf. First Bus North, 2020) indicates that this may be a feature of the West Yorkshire accent that speakers are aware of and can consciously control. If speakers in this region are aware of this stereotype, it could be the case that they use this particular vowel to signal their affiliation with West Yorkshire or alternatively, the specific borough they are from. Furthermore, people who feel particularly proud to be from the area might be more likely to use a short open-mid monophthongal FACE variant such as [ɛ]. One way in which this theory can be explored is by referring to the participants' self-evaluations of their regional identity.

As part of the WYRED data collection process, participants were asked to complete a survey in which one of the questions that they were asked to answer was "which phrase do you identify with the most?". Figure 4.13 shows the number of responses for each of the identity labels that participants could select, colour coded based on the borough that the participants were from. In this figure, the identity label 'Local' includes the names of the three boroughs as well as the town of Huddersfield. Interestingly, none of the participants from the borough of Kirklees chose the term *Kirklees*, however, three participants from this borough selected the more specific area of Huddersfield. Overall, 43% said they identified as *British* or *English* whereas 57% opted for a more local term (*Yorkshire/West Yorkshire* or a specific borough).

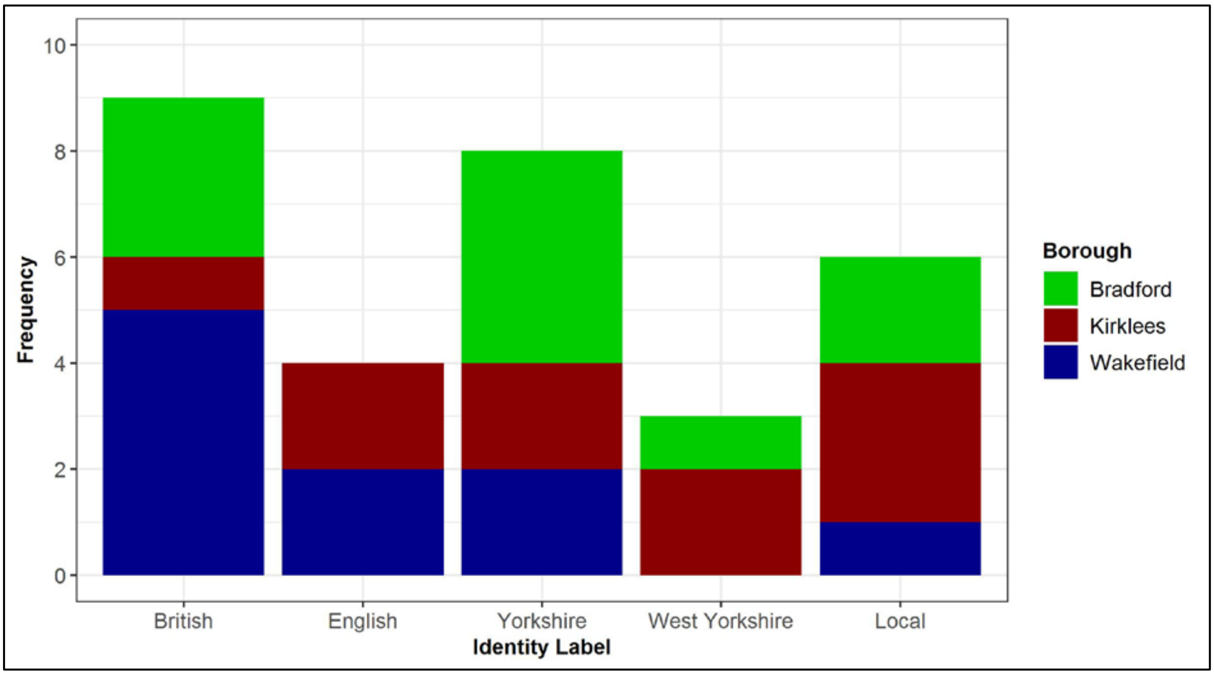


Figure 4.13. Participants' self-evaluations of regional identity by the borough they are from.

As FACE F2 values were most variable across West Yorkshire, the participants' F2 values were examined in order to observe whether or not they varied according to the participants' self-evaluations of identity. The violin plot in Figure 4.14 displays the participants' FACE F2 values grouped by the type of identity label they selected. On the left, the F2 values of participants with national identities of *British* and *English* are displayed. In the centre, the data from those who selected regional identities of *Yorkshire* and *West Yorkshire* are presented. On the right, the data from those participants who identified more locally by borough are presented. Overall, it can be seen that the distributions of F2 values vary slightly across groups, with those who have a national identity having slightly more front FACE productions than the other two groups. Based on this, it is possible that those with a national identity might realise FACE as [eː] or [eɪ], whereas those with a more local identity may be more likely to realise FACE as [ɛː] or [ɛ]. These results suggest that the participants' FACE F2 values may be closely linked to their self-evaluations of identity in addition to the borough that they are from.

Figure 4.14. Distribution of F2 midpoint formant data across self-evaluated identity types.

The data and findings reported in this chapter could be built upon to conduct various sociophonetic research projects in the future. For instance, the descriptions of FACE presented in this chapter can be compared with descriptions of FACE in other nearby towns and cities in order to assess the extent to which this vowel is regionally stratified across the North of England. Previous research has indicated that the FACE vowel is particularly variable across larger, distinct regions in the UK but the evidence presented in this investigation suggests that FACE is also variable at a local level. The fact that acoustic differences have been observed between neighbouring speech communities in this study suggests that speakers from West Yorkshire may have a particularly strong sense of local identity even at the level of their metropolitan borough. In fact, anecdotal evidence observed during the data collection process suggests that regional variation could be even more fine-grained than at the local level already observed. During the WYRED recording sessions, numerous participants claimed to be able to tell the difference between accents within the same borough (e.g. Pontefract and Hemsworth within the borough of Wakefield); however, further analyses on an even more "microscopic" level would be required to examine whether this can be corroborated with acoustic information.

The study presented in this chapter could also be replicated using speakers from a range of different sociodemographic groups across West Yorkshire, in attempt to determine how FACE varies between different social groups. In this investigation, significant differences in FACE F2 values were observed across boroughs despite that fact that the participants formed a fairly homogeneous group insofar as their social characteristics (such as sex, age group and influence of other languages). If we were to take a random sample of the male population of West Yorkshire, it is possible that the extent to which FACE varies across the region would be even greater than this. Conversely, it could be the case that other social factors have a greater effect on FACE variation and therefore region may not play as much of an important role in how FACE is constrained once other social factors are introduced. Baranowski & Turton have proposed that social class "usually turns out to be the primary source of linguistic differentiation" within any given area (2015, p. 313) and therefore it would be particularly interesting to see whether FACE is realised differently by members of other social classes within West Yorkshire. Finally, the methods of this study could be replicated in the future to assess how FACE might change over time, either by becoming more similar to surrounding areas or more distinct.

4.7. Conclusion

This chapter has revealed how the FACE vowel is realised across the metropolitan boroughs of Bradford, Kirklees and Wakefield by providing descriptions of FACE in the form of auditory transcriptions as well as vowel formant data. For each of the three boroughs, and West Yorkshire as a whole, summaries of the average midpoint F1~F3 values have been presented in addition to distance measures from the average FACE vowel onset to vowel offset. Interquartile ranges have also been provided in the form of tables and boxplots to illustrate the distribution and variability of the midpoint formant values. Using multiple linear mixed effects analyses, it has been possible to determine that acoustic differences exist between FACE realisations across the three West Yorkshire boroughs. These analyses also showed that the vowel quality of FACE is significantly different in terms of F2 across Bradford, Kirklees and Wakefield, whereas F1 and F3 were not significantly affected by regional variation. These results suggest that West Yorkshire, or Yorkshire more generally, is more regionally stratified than previously recognised.

As the primary aim of this study was to investigate how FACE was produced across the boroughs of Bradford, Kirklees and Wakefield, this chapter has mainly focused on between-speaker variation (i.e. how FACE realisations varied between participants from each of the three boroughs). Consequently, this investigation paid less attention to within-speaker variation and only briefly considered how FACE productions varied within a speaker, across tasks and across different phonetic environments, based on the results of the linear mixed effects analyses. The next chapter will systematically explore how FACE is realised in different stylistic contexts and will consider how susceptible this phonetic feature is to the influence of speech accommodation.

5. Phonetic accommodation in the West Yorkshire FACE vowel

5.1. Introduction

This chapter investigates the degree to which the FACE vowel is influenced by phonetic accommodation and explores how much within-speaker variability is present across a range of forensically-relevant scenarios[3]. The study uses semi-spontaneous conversational data and takes an acoustic-phonetic approach to evaluate levels of convergence and divergence. The realisations of FACE across three WYRED tasks within the speech of 30 West Yorkshire males and their respective interlocutors are analysed. With regards to FACE in West Yorkshire, it has been established in the study presented in Chapter 4 that subtle differences in vowel quality exist across the three boroughs of Bradford, Kirklees and Wakefield. It has also been shown that there were significant differences in FACE productions across tasks for the group of West Yorkshire speakers as a whole. However, what has yet to be determined is the extent to which FACE productions vary within each participant according to the influence of the interlocutor, and mismatch in speaking style across varying contexts.

This chapter is divided into 7 sections. The remainder of this section presents a brief summary of findings from a selection of previous speech accommodation studies that are considered to be directly relevant to this particular investigation. For a more general overview of speech accommodation research methods and findings, please refer to Section 2.1. The research questions and hypotheses for this investigation are set out in Section 5.2, and Section 5.3 provides further details about the WYRED data used in this thesis, specifically in terms of how it was used to analyse accommodation. Section 5.4 describes each of the separate methods of analysis that were incorporated into this investigation of speech accommodation, while Section 5.5 provides the results and answers each of the research questions. In Section 5.6, the results of this study are discussed and their implications are highlighted, before the conclusions of this study are given in Section 5.7.

[3] Some elements of the investigation presented in this chapter are to be published in Earnshaw (forthcoming).

5.1.1. Background research

As previously discussed in Chapter 2, phonetic accommodation has been shown to occur in a wide range of segmental and supra-segmental speech parameters, both in speech elicited in non-interactive laboratory tasks and also in conversational interactions. It seems intuitive that certain parameters may be more likely to be affected by speech accommodation than others. For instance, variables which are known to vary within a speech community depending on speaking style or social factors such as a speaker's social class, ethnicity or educational background, may be more likely to be adapted in the process of speech accommodation than parameters which are not normally particularly variable within a given community.

Labov (1972) famously made a distinction between *stereotypes*, *markers* and *indicators*, which all describe features associated with a particular dialect but vary in terms of speaker awareness. Stereotypes are widely recognised linguistic features that speakers often talk about and tend to be the subject of dialect performances and impersonations (Meyerhoff, 2011, p. 26). Whereas markers are variables that speakers are slightly less aware of but still tend to be subject to both social and stylistic variation. In contrast to this, indicators are below the level of consciousness and generally show limited or no style-shifting (Meyerhoff, 2011, p. 26). Trudgill (1986) asserted that the higher level of awareness associated with a marker, in comparison to an indicator, can lead speakers to modify their pronunciation of this variable in situations where they are paying more attention to their speech (1986, p. 10). Based on this, Trudgill suggested that in situations where people communicate with speakers of other language varieties, they are more likely to modify socially salient linguistic variables, i.e. those features of their own varieties of which they are most aware (1986, p. 11). This seems to be a logical hypothesis as speakers are likely to be aware of a wider variety of possible realisations for socially salient variables, and therefore more potential realisations are available to them.

A number of sociolinguistic studies which have examined speech accommodation across multiple speech parameters have found that convergence is more common for socially salient variables. For instance, Smith & Holmes-Elliott (2015) examined how speakers from Buckie (Northeast Scotland) adapted their speech during conversations with 'community insiders' and ‘community outsiders’ and they found that when speakers were interacting with the community outsider, there was a shift to standard variants exclusively with variables which are socially salient and/or stigmatised in the community. Trudgill (1986) reported that in his (1974) sociolinguistic investigation of Norwich English he accommodated towards his interviewees with respect to glottal variants of /t/ (a socially salient feature, classified as a marker in Norwich English); however, he did not accommodate in terms of the degree of fronting or backing in the /ɑ/ vowel (an indicator in this variety of English, considered to be below the level of consciousness). An investigation into speech accommodation carried out by Cao (2018) also indicated that socially salient linguistic variables may be more likely to evoke phonetic convergence than variables that are below the level of consciousness. Cao (2018) found that the Hong Kong English (HKE) speakers in her study tended to converge towards native British and American English model talkers on the linguistic features which were most salient to them. In this study, salient variables were defined as “sounds which have a greater phonetic difference between the HKE speaker’s native repertoire and the native interlocutor’s repertoire, and sounds which carry specific social meanings” (2018, p. 245). Further details of this study are discussed later in this section.

The findings of the above sociolinguistic studies provide motivation for examining phonetic accommodation in the West Yorkshire FACE vowel. The analysis reported in Chapter 4 demonstrated that local level variation is present in West Yorkshire, particularly in terms of the front/back dimension of the FACE vowel. Furthermore, external evidence indicates that West Yorkshire speakers are aware of this variable; such as the aforementioned First Group’s First Bus App advertisements which make use of the stereotypical monophthongal FACE variant in the word *take* by using the slogan “Want easy travel? Tech the bus” (First Bus North, 2020). The fact that this particular monophthongal FACE variant is used to index a West Yorkshire identity suggests that this variable is socially salient (i.e. speakers from this region are aware of this being a recognisable feature of the West Yorkshire accent) and therefore their choice of FACE variant might be something that they consciously control depending on the context. It is also possible that it might be heavily influenced by accommodation. For this reason, it is anticipated that this variable could be of interest to investigate for instances of phonetic accommodation.

Furthermore, a number of previous accommodation studies that have focussed on vowels have demonstrated that speakers can accommodate in terms of vowel quality. For example, Cao's (2018) investigation of short-term accommodation between native and non-native speakers of English revealed that HKE speakers adapted a range of segmental features whilst completing map tasks with British and American English model talkers. The THOUGHT and PATH vowels were analysed acoustically, as well as three consonantal parameters, and it was found that the HKE speakers diverged away from their interlocutors in terms of PATH and maintained their original difference in THOUGHT vowels. Participants also converged in terms of degree of rhoticity and their pronunciation of [z] but diverged in their pronunciation of [θ]. This highlights the fact that speakers may converge in some elements of their speech whilst simultaneously diverging in others.

Another sociolinguistic study which provided evidence of speakers modifying their vowel productions, was an investigation of long-term accommodation in speakers from a small market town in the Midlands who had moved away to university (Evans & Iverson, 2007). The aim of this study was to examine how young adults adapted their accent as a consequence of their educational setting and interacting with speakers of SSBE. Participants were recorded reading aloud a list of sentences and a reading passage on four separate occasions: once prior to beginning university, three months later and then at the end of their first and second years of study. It was determined that participants changed their production of a range of vowel sounds, in some cases becoming more similar in their pronunciation to that of a SSBE accent. For example, by analysing their F1 and F2 values, it could be seen that the vowels in the words *bud* and *cud* became more fronted and lower over time.

Babel (2009) investigated short-term phonetic imitation of vowels in California English by analysing the speech of over 150 participants undertaking a lexical shadowing task. In this sociophonetic study, it was found that participants converged towards the two model talkers in terms of F1 and F2. Convergence rates were found to be inconsistent across vowel categories, with low vowels exhibiting stronger imitation effects than high vowels. Interestingly, the findings of this study went against Babel's prediction that, following Trudgill (1981), the vowels undergoing a sound change in California English, /o/ and /u/, would "exhibit the strongest effects of imitation because participants have a wider variety of stored representations" for these vowels (Babel, 2009, p. 57). In a later publication related to this study, Babel suggested that the reason that the low vowels were imitated more consistently and to a greater extent than the high vowels may have been to do with the difference in dialect backgrounds between the participants and the model talkers (2012, p. 186). Specifically, it was hypothesised that participants may have converged more strongly towards the model talker's low vowels because their own low vowels were far away from the model talker's to begin with, and therefore participants had "acoustic-phonetic space in which to accommodate" (2012, p. 186). A similar hypothesis was put forward in Cao (2018) where it was observed that rates of convergence were higher for linguistic variables that were easier for the participants to perceive as being different in their own speech from the model talkers.

Based on the findings reported above, we might predict that the participants in the present study will display higher rates of convergence during the mock police interview task, where they interact with someone with a different accent to their own, than in the casual paired conversation. On the other hand, in Pickering & Garrod's (2013) account of language production and comprehension, they describe a "simulation route" whereby "comprehenders will emphasize simulation when they are similar to the speaker because simulation will tend to be accurate" (2013, p. 18). Also, in line with this theory, Pardo et al. (2018b) have proposed that "talkers who are more similar to each other should converge more than those who are less similar, so that same-sex pairs of talkers should converge more than mixed-sex pairs." (2018b, p. 4). Based on these accounts, we might instead expect to find higher levels of convergence in the paired conversation than in the mock police interview pairings, due to the fact that in the paired conversation the participants are matched in terms of age, gender and where they were from and were therefore likely to have relatively similar FACE productions.

It should be highlighted that the data used in most of the aforementioned studies was elicited in slightly contrived settings (i.e. speech-shadowing, reading and map tasks) and the extent to which these findings would be replicated in more natural conversational data is unknown. The current study aims to investigate how much speech accommodation occurs in FACE vowel productions when the data is elicited in a context which is perhaps more closely aligned with naturally occurring spontaneous speech. A broader aim of this research is to establish what influence speech accommodation could have on evidential speech samples being analysed for FSC casework, and therefore, by using forensically-relevant spontaneous speech, it is anticipated that the findings of this investigation will be applicable to real-world scenarios.

In Chapters 6 and 7, word-medial, intervocalic /t/ will be analysed within the speech of the same 30 West Yorkshire participants. The findings of the four case studies presented in this thesis will subsequently be evaluated in order to determine which, if any, of these two speech parameters are most affected by the influence of speech accommodation. Based on the findings above, it is anticipated that the participants will accommodate most with respect to linguistic variables that are considered to be socially salient in West Yorkshire.

5.2. Research questions and hypotheses

The study presented in this chapter aims to establish the extent to which speakers from West Yorkshire accommodate in their productions of the FACE vowel, across a range of forensically-relevant scenarios. In order to do this, the following research questions are addressed:

1. What is the influence of the task on FACE productions?
2. How consistent are FACE productions within tasks by speaker?
3. How do the participants' FACE productions change as a result of exposure to their interlocutor?
    a. Do participants accommodate during the paired tasks?
    b. How does accommodation behaviour across the two paired tasks vary?
    c. Is accommodation behaviour correlated across the two paired tasks?
    d. Does accommodation behaviour vary over the course of each paired task?

The first research question has been partially addressed in Chapter 4, where it was shown that there were statistically significant acoustic differences in FACE productions across the three speaking tasks considered in this study. In this chapter, this finding is explored in more detail and further quantitative information is provided to illustrate exactly how FACE productions varied across tasks, as a potential consequence of the interlocutor and the varying speaking contexts. It is hypothesised that formant values will vary across tasks in terms of average values at the individual level and at the group level. It is worth reiterating here that in this study, accommodation is defined broadly to refer to adaptations in speech across situations involving different interlocutors and speaking styles. Although it is not possible to separate the effects of speaking style from those of the interlocutor, the findings of this research question are still valuable for establishing the degree to which FACE productions vary across forensically-relevant speaking tasks.

With regards to the second research question, consistency is examined with reference to standard deviations of FACE formant values within three speaking tasks. It is predicted that FACE productions will become less consistent when participants are interacting with an interlocutor, as a consequence of accommodation, and therefore standard deviations are expected to be lowest during the baseline task, where there is no interlocutor present.

To address the third research question, and each of its sub-parts, the influence of exposure to the interlocutor is evaluated by measuring how similar participants' FACE productions are to their respective interlocutors. This is undertaken by calculating the Euclidean distances between interlocutor pairs at various points in the investigation. Firstly, distances between interlocutor pairs whilst interacting in the paired tasks are compared to distances between the participants in their baseline task and their partner in the paired task. The aim of this assessment is to explore to what extent participants accommodate whilst interacting with an interlocutor. Secondly, accommodation behaviour is compared across two paired tasks involving different interlocutors in order to evaluate in which scenario participants are more likely to accommodate and it is also considered whether or not the results are correlated across tasks. Finally, changes in distances between interlocutor pairs over the course of a paired task are tracked. The purpose of this assessment is to establish whether there is an increase in any accommodation behaviour that occurs during a paired task, as a result of more exposure to the interlocutor over the course of the interaction.

Based on the findings of the studies discussed in Section 5.1.1, it is predicted that the participants in the present study will adapt their FACE productions according to whom they are talking within each of the paired tasks. As the FACE vowel is considered to be socially salient in West Yorkshire, it can be expected that some form of speech accommodation will take place, which can be observed quantitatively by way of measuring the FACE vowel formants. In relation to changes over the course of an interaction, it is anticipated that participants who converge towards their interlocutor will become more similar to them following more exposure. For participants who diverge, it is expected that divergence will take place early on in the interaction and remain relatively consistent over the course of the task.

It is also predicted that participants will accommodate to different degrees, and potentially in different directions, across the two (separate) paired tasks. This hypothesis is based on the fact that the paired tasks differ in many ways; including the levels of formality required, same/mixed-sex pairings, how similar the accents of the interlocutor pairs are to begin with, and the amount of pressure the participants feel under due to the differing power dynamics at play. Due to the spontaneous nature of the conversational data analysed in this study, all of the above factors need to be taken into account when considering how the participants might adapt their FACE productions across tasks. As is the case with most empirical research involving spontaneous speech, it is not possible to separate the effects of different speaking styles across tasks from the effects of the interlocutors in this study. This is an even more complicated task in real-world scenarios such as FSC casework involving authentic recordings made in contexts involving different speaking styles and interlocutors.

By considering some of the differences between the two paired tasks, it is possible to make predictions about how the participants' behaviour might change across tasks. For instance, it has previously been suggested that accommodation studies involving collaborative tasks or games might elicit higher levels of convergence compared to casual conversations (Levitan & Hirschberg, 2011; Schweitzer & Lewandowski, 2014), therefore we might expect to find higher levels of convergence in the mock police interview task than in the casual paired conversation, due to the fact that the task is collaborative in nature, with the police officer asking questions and the participants having to find clues to the answers on an iPad. On the other hand, this task is potentially more stressful and the fact that the participants have to lie in some instances might cause the participants to react negatively towards the police interviewer. It has been shown that positive feelings towards an interlocutor can evoke higher rates of convergence (Babel, 2010; Byrne, 1971) whereas negative feelings can lead to divergence as a means to express distain (Bourhis & Giles, 1977; Shepard et al., 2001). Therefore it may be the case that participants diverge away from the police interviewer.

As the mock police interview task involves the participants interacting with someone from outside of West Yorkshire, greater baseline differences are expected between pairs of interlocutors in this task than in the paired conversation task where participants interacted with someone from the same area. However, it is acknowledged that previous studies have reported mixed results in relation to the effects of the existing phonetic repertoire on phonetic convergence and therefore it is unclear whether greater baseline differences will lead to higher or lower levels of convergence. Furthermore, it is not known how the multiple factors mentioned above will interact with one another, and so it is difficult to predict how participants in this study will accommodate in the two paired tasks.

5.3. Data

This study analysed the FACE tokens from 30 participants produced during WYRED Tasks 1, 3 and 4 that were examined in the previous investigation of how FACE is realised across West Yorkshire (in Chapter 4). Tasks 1 and 3 were paired tasks in which the participants interacted with another person, with the former involving a mock police interview scenario and the latter involving a casual conversation. Task 4 consisted of participants leaving an answer message for their fictional brother asking for help in relation to the crime scenario that had been discussed during the mock police interview task. Figure 5.1 provides a schematic of the three WYRED tasks that each participant completed, illustrating who the participants interacted with in the paired tasks and the one-way communication in Task 4.

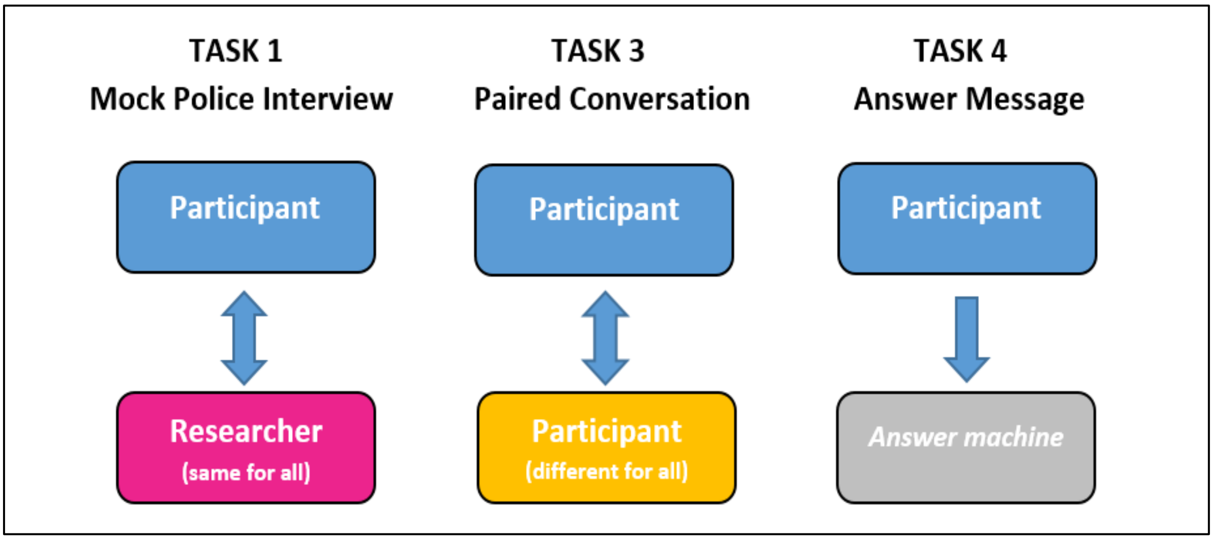


Figure 5.1. Schematic of the three WYRED tasks.

Only the midpoint F1~F3 FACE vowel data were used in this investigation, as it was determined in the previous chapter that the F1 and F2 distance measures from FACE vowel onset to offset did not significantly vary across boroughs or tasks. Furthermore, the majority of tokens were monophthongal and therefore midpoint formant values were considered to adequately represent the realisations. Table 5.1 provides a summary of the number of FACE tokens produced by the West Yorkshire participants included in this study, across the three tasks. There were considerably less tokens available in Task 4 than in Tasks 1 and 3 due to the relatively short length of this task (average length of 2 minutes versus 25 and 21 minutes, respectively).

Table 5.1. Summary of the participants' FACE tokens per task.

Task	Total	Average per participant	Median per participant
1	967	32	34
3	881	29	31
4	268	9	9

For Tasks 1 and 3, the FACE tokens of each participant and their interlocutor were analysed in order to draw conclusions about how the interlocutors' speech may have influenced the participants' FACE productions. As each participant was paired with another participant from the same area as themselves in the Task 3 recordings, FACE tokens were already segmented and analysed acoustically for both interlocutors in each pair. However, for the Task 1 recordings, participants spoke to a female researcher from Gateshead and therefore her FACE tokens needed to be segmented and analysed separately. The procedure outlined in Section 4.4.2 was followed in order to measure up to 35 FACE tokens from each of the researcher's Task 1 recordings.

In total, 914 FACE tokens were extracted from the 30 Task 1 recordings (average of 30 tokens per recording) and midpoint F1~F3 values were measured and logged. The researcher's FACE vowel realisations mainly varied between the diphthongal [eɪ] and [ɛɪ] variants and open-mid monophthongal variants in the vicinity of [ɛː] and [eː], depending on the phonetic context of the token. Theoretically, West Yorkshire participants producing more diphthongal realisations or variants with higher F1 and F2 values would be taken to indicate convergence towards the researcher. However, based on auditory impressions documented in the previous chapter, the participants rarely tended to produce diphthongal [eɪ] variants across any of the tasks.

In order to identify and assess any potential speech accommodation that may have taken place, FACE tokens produced during Task 4 were compared to tokens from the two paired tasks. By including two paired tasks, with different interlocutors in each, it was possible to consider the effects of differing "model talkers" on speech accommodation. As well as considering how FACE varies according to interlocutor, by examining how FACE is realised across the baseline (Task 4) and paired tasks (Tasks 1 and 3), the effects of speech style and context also contributed to the findings of this investigation. As the nature of the task changed between each of the participant's recordings, it was not possible to disentangle the influence of the interlocutor from other contextual factors, such as the purpose of the conversation, varying speech styles and levels of formality required, and whether or not the participant was talking over the telephone. While this made the investigation into speech accommodation more challenging, it also made the findings of this study more forensically relevant, as authentic samples examined in FSC casework are often made in contexts involving different speaking styles and interlocutors and are typically mismatched in terms of channel. Using the WYRED Task 1, 3 and 4 data, it was anticipated that the full extent to which the FACE vowel varies within an individual could be assessed.

5.4. Methodology

This section outlines the methods employed to examine phonetic accommodation in the West Yorkshire FACE vowel. It describes three different methods for assessing the variability in FACE productions and provides details of transformation techniques that were applied in this study.

5.4.1. Measuring the influence of the task on FACE

In Chapter 4, a series of linear mixed effects analyses were presented which explored the effects of BOROUGH, TASK and PHONETIC ENVIRONMENT on the midpoint F1~F3 values of all participants' FACE tokens. The linear mixed effects model with midpoint F1 values as the dependent variable which was used to perform this analysis is shown again below:

F1.midpoint.model = lmer(F1 ~ BOROUGH + TASK + ENVIRONMENT + (1+TASK|PARTICIPANT), data=Data, REML=FALSE)

This chapter reiterates the findings of these analyses with respect to the effect of TASK and discusses whether there were any overwhelming patterns across the participants, whereby one task elicited significantly higher formant values than another. As well as evaluating if significant differences existed between tasks for the group of participants as a whole, it further explores how FACE varied across tasks in more detail by establishing the average midpoint F1~F3 values for each participant's FACE tokens across the three tasks. These average values were subsequently plotted in order to visualise the amount of change across tasks as well as to illustrate the range of movement in FACE productions across tasks.

5.4.2. Measuring the consistency of FACE within tasks

In addition to evaluating how FACE changed across tasks, an assessment was made of how consistent the FACE vowel productions were for each participant within each task (i.e. the degree to which the midpoint formant values of FACE varied across tokens). To quantify the levels of variation of FACE, the standard deviation (SD) of each participant's formant values for F1~F3 was calculated. The SD expresses how much each of the formant values differ from the average of the full set of FACE formant values. For each participant, a separate SD was calculated for each of the three tasks that each participant undertook. Therefore, the SD values captured how much each of the midpoint values from a specific task differed from the average formant value for the task in question. These values were subsequently compared between tasks for each participant to examine in which task the participants displayed the highest levels of variation. Higher SDs were taken to indicate less consistent FACE vowel realisations within a task.

Comparisons across tasks at the group level were also carried out using a series of linear mixed effects analyses in which the fixed effects of BOROUGH and TASK (without interaction terms) were used to predict the SDs of each formant. As random effects, there were intercepts for PARTICIPANTS. The full linear mixed effects model with the F1 SD values as the dependent variable is presented below:

F1.SD.model = lmer(F1.SD ~ BOROUGH + TASK + (1|PARTICIPANT), data=SD.Data, REML=FALSE)

This model was subsequently re-run using F2 and F3 SD values as the dependent variable. All models described in this chapter were fitted using maximum likelihood and visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. In order to test the influence of each of the fixed effects on the F1~F3 SD values, respectively, p-values were obtained by likelihood ratio tests of the full model against the model without each of the fixed effects in question. It should be noted that, as with the analysis of regional variation in the FACE vowel in Chapter 4, it was not deemed necessary to use transformed formant values before calculating the SDs because all participants were of the same biological sex and were of similar ages. Raw formant values were therefore used to calculate the averages and SDs of each participant's FACE productions across tasks and to run the linear mixed effects analyses.

It was hypothesised that there would be higher levels of variation in cases where participants were found to adapt their speech depending on who they were talking to, than when participants did not accommodate. If higher SDs were reported in the paired tasks than in Task 4, this was taken to indicate that some form of speech accommodation had taken place over the course of the interaction; whether that be convergence towards the interlocutor or divergence away from the interlocutor. Whereas, low SD values in the paired tasks indicated that FACE vowel realisations were relatively consistent across the duration of the task and therefore this phonetic parameter did not appear to be changing as a result of more exposure to the interlocutor. This theory would be supported if further analysis of formant values revealed consistent changes over the course of the interaction. It was anticipated that overall the participants' FACE productions would be most consistent in Task 4, where there was no interlocutor providing feedback to the participants.

One confounding factor that needs to be acknowledged, however, is the uneven number of FACE tokens per task across participants. As was reported in Section 5.3, there were considerably less FACE tokens available for analysis in Task 4 than in Tasks 1 and 3, due to the fact that Task 4 was a much shorter task than the two paired tasks. There were only 9 tokens on average in the Task 4 recordings, compared with approximately 30 tokens in the paired tasks. As time was limited in Task 4, this may have allowed for less variation to occur resulting in lower SD values. Furthermore, the phonetic environment of FACE has been shown to influence productions and therefore a participant who produced FACE tokens from a wide range of phonetic environments in Task 1 but not in Task 4 might be expected to have higher SD values in Task 1 than Task 4, as a result. The relatively low number of tokens available in the Task 4 recordings also meant that the averages and SDs calculated for FACE tokens in this task may be more heavily influenced by outliers than those calculated for Tasks 1 and 3 and therefore these values must be treated with some degree of caution. Unfortunately, this was not something that could be controlled due to the spontaneous nature of the speech considered in this investigation.

5.4.3. Assessing the influence of the interlocutor on FACE

The first two methods of analysis described above are intended to shed light on the degree to which speakers vary their FACE vowel productions within tasks and between tasks. While this is useful for assessing levels of within-speaker variation, it does not provide sufficient information to evaluate speech accommodation thoroughly. Although inferences can be made that speech accommodation may have occurred in instances where high levels of variation are present within and between tasks, it could well be the case that this variation is not related to the interlocutors' speech but rather the context of the task in terms of speech style or topic, for example. In order to conduct a thorough investigation which adequately addresses the question of how much the participants accommodate in terms of their FACE vowel, it is necessary to explore the influence of the interlocutor by examining how their realisations of the FACE vowel relate to the participants' productions. This section of the analysis therefore aims to determine how much participants converged towards, or diverged away from, their interlocutors during the paired tasks (Task 1 and Task 3). It should, however, be reiterated here that even with these separate approaches it is not possible to fully separate the effects of speaking style and the interlocutor from one another.

Euclidean distances within the F1 x F2 vowel space between interlocutors were measured at various points over the course of the study in order to explore the extent to which participants accommodated whilst interacting with an interlocutor. Distances between interlocutor pairs were anticipated to be greater in Task 1 than in Task 3, because the interlocutor in Task 1 was female and therefore her raw formant values were expected to be much higher than those of the male participants as a consequence of anatomical differences. For this reason, all formant data across all tasks was transformed prior to calculating Euclidean distances. The next sub-section describes the method that was used to transform the raw formant values, followed by details of the procedure that was used to calculate Euclidean distances and quantify accommodation more generally in this investigation.

5.4.3.1 Transforming the vowel formant data

Before the Euclidean distances could be calculated, it was necessary to transform all of the raw vowel formant data across all tasks. The reason for this was that distances were to be measured between all interlocutor pairs which meant that for the Task 1 recordings, distances between the vowel formant data of the male participants and the female researcher's speech needed to be calculated. As it is widely acknowledged that biological differences between men and women such as differing vocal tract lengths tends to result in females having a higher baseline f_0 compared to males, it was anticipated that the researcher's formant values would be significantly higher than those of the male participants.

Table 5.2 presents a summary of the researcher's average formant values and SDs compared to those of the male participants. Figure 5.2 also provides a visualisation of the vowel distributions in the form of an ellipse for the researcher and the participants (grouped by area). The researcher's data is represented by the orange "Gateshead" ellipse and the centre of the area labels represent the average F1 and F2 values for each area's distribution. The x- and y-axes have been reversed in line with standard conventions of vowel plots, in order to better represent the vowel space.

Table 5.2. Summary of FACE formant values for West Yorkshire participants and the researcher.

		Formant values (Hz)		
	Measure	**F1**	**F2**	**F3**
West Yorkshire Participants (N=2116)	**Average**	547	1747	2516
	SD	55	164	172
Researcher (N=914)	**Average**	676	2170	2952
	SD	61	150	210

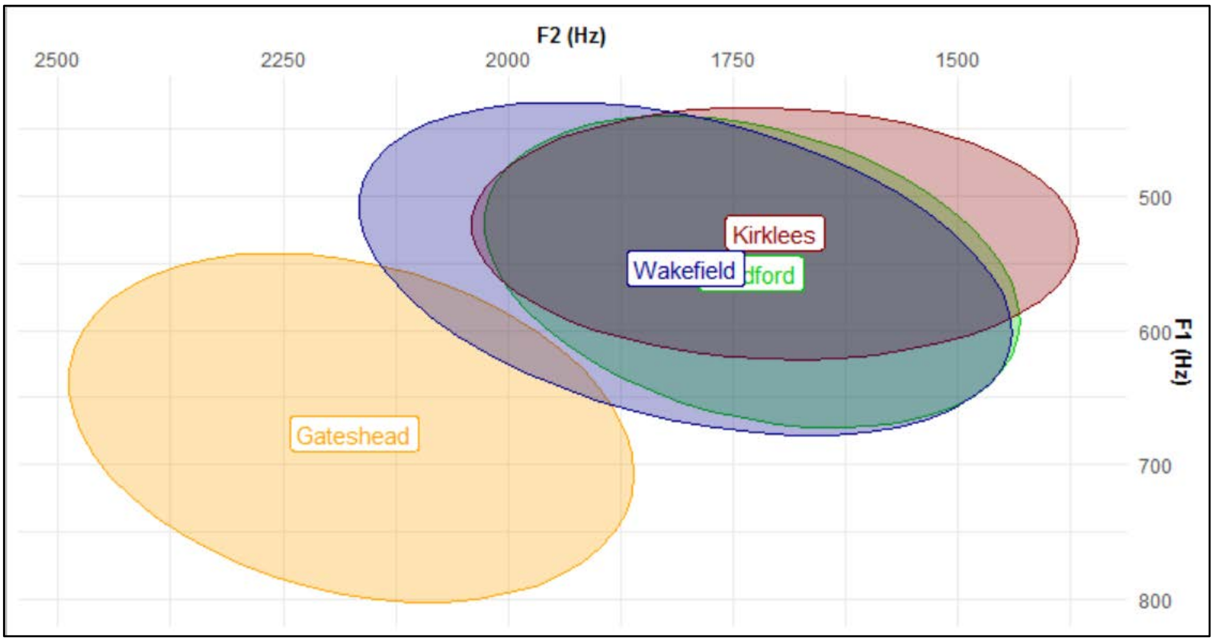


Figure 5.2. Vowel plot showing the distribution of the raw FACE formant values of all participants and the researcher.

In Figure 5.2, it can be seen that there is a stark difference between the formant values of the researcher's FACE vowels compared to the participants' FACE vowels. Consequently, it was expected that much greater distances would be present in the Task 1 recordings than in the Task 3 recordings if raw formant values were used to calculate Euclidean distances. In order to ensure that any conclusions relating to speech accommodation were comparable between

Task 1 and Task 3 and were unaffected by factors such as vocal tract length, it was necessary to transform the raw formant data.

There are a wide range of different vowel normalisation and transformation procedures that have been used for sociolinguistic research previously (for a review of 20 different methods see Flynn (2011)). The Bark Difference Metric (BDM) method, which was modified from the formula developed by Syrdal & Gopal (1986), was considered to be the most suitable vowel transformation procedure for the current study. One reason for this was that this method is vowel-intrinsic (i.e. it only requires data from one vowel category) which was crucial for this study as formants had only be measured for the FACE vowel, rather than the whole vowel space. This method is also speaker-intrinsic which means that it uses information from a single speaker to perform the transformation, rather than calculating a single average value for all speakers. This was important because speaker-extrinsic normalisation methods generally require a very high number of speakers (approximately 350) in order to perform effectively, and without this they have been found to introduce distortions of some normalised values (Thomas, Kendall, Yaeger-Dror, & Kretzschmar, 2007). As the current study only included 31 speakers, it was unlikely that a speaker-extrinsic method would perform well. The BDM procedure is also formant-extrinsic which means that it uses information from multiple formants to transform the formant values. Thomas et al. (2007) note that the biggest disadvantage of this method is that it is heavily dependent on F3. This can be a problem in cases where the audio is of poor quality and it is not possible to obtain reliable F3 values; however, this was not a concern for the current study as the formant data was all from high-quality studio recordings. Perhaps most importantly, Thomas et al. (2007) state that this method is able to filter out physiological differences while retaining sociolinguistic differences, which was the main aim of transforming the data.

The formant data was transformed using the NORM vowel normalisation and plotting suite v1.1 (Thomas et al., 2007). To implement the BDM transformation method, NORM converts all of the individual F1~F3 formant values from Hertz (Hz) to Barks (Z) using the equation given by (Traunmüller, 1997), presented below:

$$Z_i = 26.81/(1+1960/F_i) - 0.53$$, (Where F_i is the value for a given formant *i*)

Once the formant values had been converted into Barks, the differences Z3 - Z1, Z3 - Z2, and Z2 - Z1 were computed in NORM. Z3 - Z2 was used to plot the transformed front-back dimension and Z3 - Z1 was used to plot the transformed height dimension. From this point onwards, *transformed F1* refers to the Z3 - Z1 bark values, *transformed F2* refers to Z3-Z2 bark values and *transformed F3* refers to Z2-Z1. Table 5.3 presents a summary of the researcher's BDM transformed average formant values and SDs compared to those of the male participants. Figure 5.3 also presents a visualisation of the distribution of the researcher's FACE vowel data and that of the participants, using the BDM transformed formant values. Again, the participants' data has been grouped according to the borough that they are from (Bradford, Kirklees and Wakefield). The researcher's data is represented by the orange "Gateshead" ellipse and the centre of the area labels represent the average F1 and F2 values for each area's transformed distribution. It can be seen that there is now much more overlap between the data of the participants and the researcher; however, some of the differences are still preserved as we would expect due to the fact that the researcher is from Gateshead and realises FACE differently to the West Yorkshire participants.

Table 5.3. Summary of BDM transformed FACE formant values for West Yorkshire participants and the researcher.

		BDM transformed formant values (z)		
	Measure	**F1**	**F2**	**F3**
West Yorkshire Participants (N=2116)	**Average**	9.21	2.44	6.77
	SD	0.64	0.55	0.86
Researcher (N=914)	**Average**	9.22	2.02	7.20
	SD	0.62	0.41	0.74

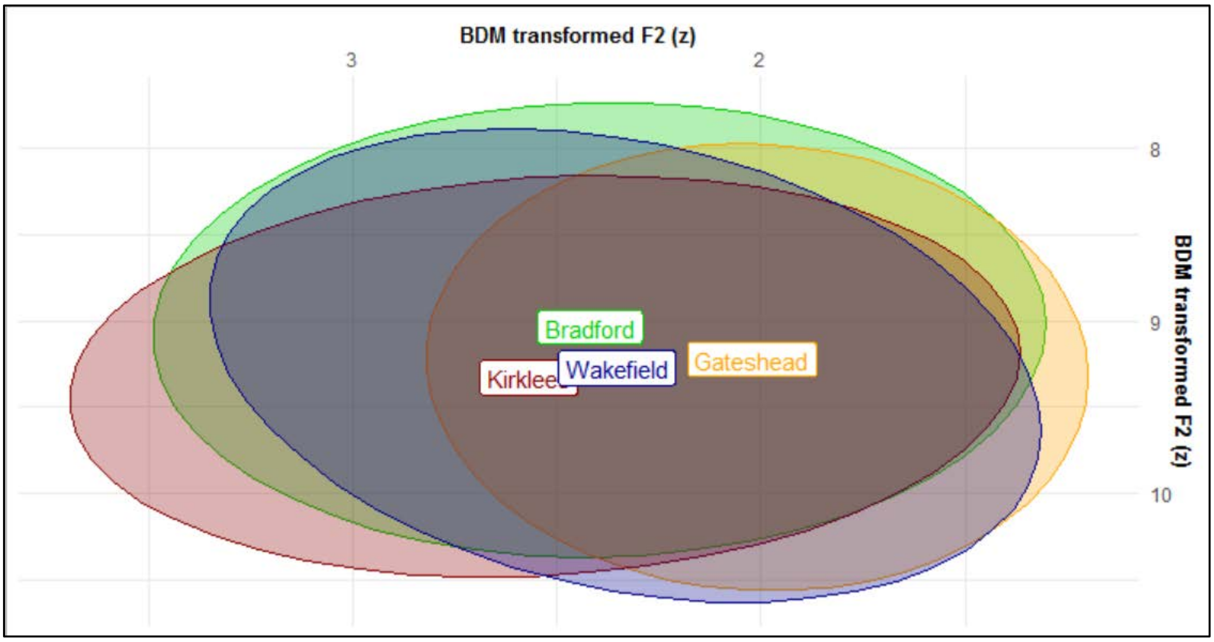


Figure 5.3. Vowel plot showing the distribution of the BDM transformed FACE formant values of all participants and the researcher.

5.4.3.2 Measuring Euclidean distances

In order to quantitatively measure how each participant's FACE realisations changed as a result of exposure to their interlocutor, Euclidean distances within each pair of interlocutors in the Task 1 and Task 3 recordings were calculated. The Euclidean distance is the straight-line distance between two points in Euclidean space. In this study, the Euclidean distance is measured within the two-dimensional vowel formant space (F1 x F2). The equation for calculating the Euclidean distance between each pair is presented and explained below.

$$\sqrt{\left(Average\ F1_{interlocutor} - Individual\ F1_{participant}\right)^2 + \left(Average\ F2_{interlocutor} - Individual\ F2_{participant}\right)^2}$$

Figure 5.4 provides a visualisation of the three speaking tasks and the vowel data that were used from each task to measure different sets of Euclidean distances. Using the BDM transformed FACE F1 and F2 values (z), the following steps were taken for each participant, first for Task 1 and then for Task 3:

1. To establish how their interlocutor typically realised FACE, average F1 and F2 FACE values were calculated.
2. To assess how similar the participant's FACE tokens were to their interlocutor's FACE realisations when they were not interacting with one another, *baseline distances* were calculated*. Baseline distances* were calculated by measuring the Euclidean distance between the average F1 and F2 FACE values for the interlocutor, and the F1 and F2 values of each individual FACE token produced in Task 4.
3. The average of the *baseline distances* was calculated in order to measure the typical difference between the participant and their interlocutor when they were not interacting with one another. This average value was referred to as the *original distance*.
4. To establish how similar the participant's FACE tokens were to their interlocutor's FACE realisations when they were talking to one another, a set of *paired distances* were calculated. *Paired distances* were calculated by measuring the Euclidean distance between the average F1 and F2 FACE values for the interlocutor and the F1 and F2 values of each individual FACE token produced in the paired task.
5. To determine the amount of change from the *original distance* between interlocutors to the *paired distance* between interlocutors, the *original distance* was subtracted from each of the *paired distances.* The resulting set of values were termed *difference in distance* values.

In cases where a *paired distance* was smaller than the *original distance*, resulting in a negative *difference in distance* (DID) value, the acoustic distance between the participant and their interlocutor had reduced. This indicated that some degree of phonetic convergence took place. Conversely, productions eliciting positive DID values were those where the distance between interlocutors during the paired task were greater than the original distance, indicating that the participant had diverged away from their interlocutor during the paired task. Theoretically, a value of 0 would indicate no change as a result of auditory exposure to the interlocutor. The DID values were therefore considered to be one of the crucial measures of speech accommodation in this study.

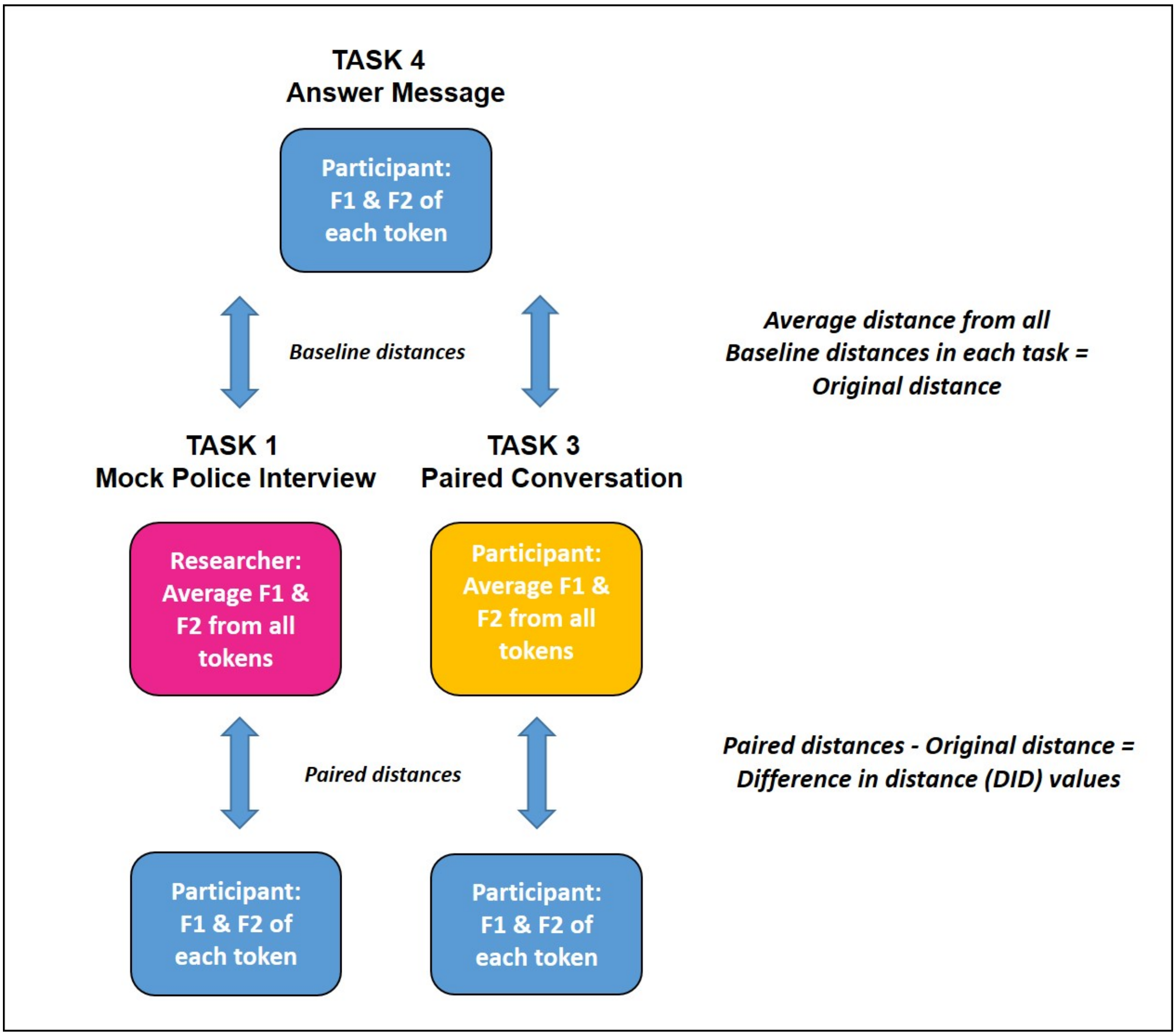


Figure 5.4. Schematic of the calculation of Euclidean distances using BDM transformed FACE F1 and F2 data from each of the WYRED tasks.

Before outlining how the DID values were used to examine accommodation, it may be useful to illustrate the above procedure with an example. For participant #064, 35 FACE tokens were measured in Task 1, and 13 tokens were measured in Task 4. 35 of the researcher's FACE tokens produced during participant #064's mock police interview were also measured and her average transformed F1 and F2 values were 9.43 z and 2.11 z, respectively. Euclidean distances were calculated between the researcher's average F1 and F2 values and the F1 and F2 values of each of the participant's FACE tokens in Task 4. This resulted in 13 baseline distances. The average of these baseline distances was then calculated to arrive at the original distance between interlocutors: 1.11 z. Subsequently, Euclidean distances were calculated between the researcher's average F1 and F2 values, and the F1 and F2 values of each of participant #064's FACE tokens in Task 1 - resulting in 35 paired distances. The original distance (1.11 z) was then subtracted from each of the paired distances in order to determine the DID of each token. The resulting DID values were then evaluated to determine levels of convergence during the paired task.

For all participants, the DID values from the two paired tasks were compared in order to establish whether participants were more likely to accommodate towards their interlocutor in Task 1 or in Task 3. To do this, a linear mixed effects analysis was performed in which the fixed effects of BOROUGH, TASK and PHONETIC ENVIRONMENT (without interaction terms) were used to predict the DID values. As random effects, there were intercepts for PARTICIPANTS as well as by-participant random slopes for the effect of TASK. In order to test the influence of each of the fixed effects on the DID values, p-values were obtained by likelihood ratio tests of the full model against the model without each of the fixed effects in question. The full linear mixed effects model with DID values as the dependent variable is presented below:

DID.model = lmer(DID ~ BOROUGH + TASK + ENVIRONMENT + (1+TASK|PARTICIPANT), data=DID.Data, REML=FALSE)

The effects of exposure to the interlocutor over time were also assessed by dividing the FACE data for each participant into two halves and comparing the DID values from the first half of FACE tokens to those from the second half, in each of the paired tasks. In cases where there were an uneven number of tokens within a task, the central token was removed. The purpose of this was to examine whether participants became more or less similar to their interlocutor following more exposure. This was tested using another linear mixed effects analysis, whereby the fixed effects of LATENCY (early or late tokens), BOROUGH, and PHONETIC ENVIRONMENT (without interaction terms) were used to predict DID values. As random effects, there were intercepts for PARTICIPANTS. The model below was first fitted using Task 1 data and then using Task 3 data:

DID.full.model = lmer(DID ~ BOROUGH + LATENCY + ENVIRONMENT + (1|PARTICIPANT), data=t1data, REML=FALSE)

5.5. Results

This section presents the results of the accommodation analysis. Section 5.5.1 summarises the results of the linear mixed effects analyses related to the differences in FACE productions between tasks. Section 5.5.2 presents statistics relating to the SD of the FACE formant values to show how much within-task variation there was. Section 5.5.3 sets out summaries of the Euclidean distances that were calculated between pairs of interlocutors and quantifies within-speaker variation, whilst also providing results of further statistical analyses.

5.5.1. What is the influence of the task on FACE productions?

FACE productions were shown to vary to a significant degree across the three tasks when considering the formant data for the 30 participants as a whole. The linear mixed effects analyses presented in Section 4.4.4 examined the relationship between the quality of the FACE vowel (midpoint formant values) and the fixed effects of BOROUGH, TASK and PHONETIC ENVIRONMENT. The p-values that were obtained by likelihood ratio tests of the full model with the fixed effect of TASK against the model without the effect of TASK showed that there were significant differences in terms of F2 values across the three tasks overall. Taking Task 4 as the intercept in the models, TASK significantly affected F2 ($\chi2(2) = 15.04$, $p < 0.001$), lowering it by about 57 Hz ± 14 (standard errors) for Task 1 and lowering it by about 60 Hz ± 13 (standard errors) for Task 3. The likelihood ratio test results showed that F1 and F3 values did not vary significantly across the three tasks. However, it was noted that the F1 model failed to converge when including by-participant random slopes for the effect of TASK. Without the random slopes, TASK significantly affected F1 ($\chi2(2) = 14.26$, $p < 0.001$), increasing it by about 10 Hz ± 3 (standard errors) for Task 1 and increasing it by about 3 Hz ± 3 (standard errors) for Task 3. Nevertheless, it was felt necessary to account for the fact that the effect of TASK may not be equal for all participants and therefore only the differences in F2 values were considered to be significantly different overall.

These findings indicated that the context of the interaction during the tasks may have influenced how the FACE vowel is realised. Results suggest that it could be the case that the F2 dimension of the FACE vowel, relating to vowel front/backness, may be most susceptible to speech accommodation. This is an interesting finding as FACE F2 values were also found to be most regionally marked across West Yorkshire. This could therefore be evidence in support of the theory that phonetic accommodation is more likely to occur in relation to features that are considered to be socially salient in the area under investigation. As mentioned in Chapter 4, it was also determined that the fixed effects of BOROUGH and PHONETIC ENVIRONMENT were significant predictors in some of the models. Full summary tables for the midpoint F1~F3 models are included in Appendix 3.

Figure 5.5 provides a visualisation of the vowel distributions in the form of an ellipse for each of the three tasks. The participants' data has been grouped according to the borough that they are from (Bradford, Kirklees and Wakefield). The reason for this is that it has been found that FACE realisations are significantly different in terms of midpoint F2 values across the three boroughs. The x- and y-axes have been reversed in line with standard conventions, in order to better represent the vowel space.

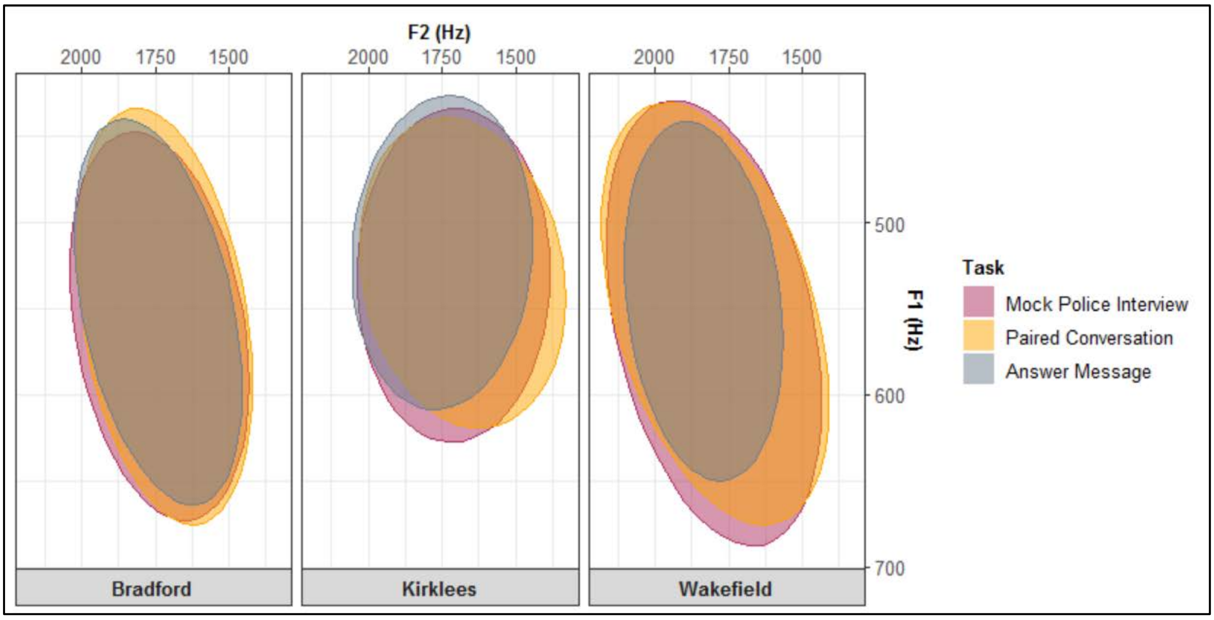


Figure 5.5. Vowel plot showing the distribution of the raw FACE formant values of all participants across tasks and West Yorkshire boroughs.

For the Bradford participants, it can be seen that overall the distributions of FACE are relatively similar across the three tasks. A larger amount of variation can be seen between tasks in the Kirklees and Wakefield data. Within the Kirklees participants, it would appear that in Task 4 the FACE F1 values are generally lower and F2 values are slightly higher than in Tasks 1 and 3.

This means that overall, FACE is produced slightly more fronted and close in Task 4 than in the two paired tasks. It also seems as though the Task 4 data is less spread in terms of F1 and F2 compared to Tasks 1 and 3 in the boroughs of Kirklees and Wakefield. This could be due to the fact that there were fewer tokens produced in Task 4 than there were in Task 1 and Task 3, but it could also indicate that the participants' FACE productions were more similar to one another during Task 4 than in the two paired tasks. It is perhaps surprising that the effect of task appears to be different across the three boroughs; however, this may be a consequence of individual variability in accommodation behaviour as opposed to being regionally stratified.

In terms of the average midpoint values, it was found that relatively little differences were observed between the three tasks. Table 5.4 presents a summary of the average midpoint F1~F3 values across Tasks 1, 3 and 4, in each of the boroughs of Bradford, Kirklees and Wakefield, as well as the three areas together. The boxplot in Figure 5.6 also visualises the average formant data for all participants across all three tasks. Please note that the y-axes for each of the F1~F3 subplots have been customised within a range the best suits the average formant values and therefore they are not all the same. By considering the values presented in Table 5.4 and the distributions of average formant values represented in Figure 5.6, it can be seen that the average formant values for FACE vary only slightly across tasks, when taking into account the formant data of all 30 participants, or when separated by borough.

Table 5.4. Average formant values for FACE across Tasks 1, 3 and 4, in the boroughs of Bradford, Kirklees, Wakefield and all three areas together.

		Average formant values (Hz)		
Area	**Task**	**F1**	**F2**	**F3**
Bradford	1	562	1741	2511
	3	555	1717	2462
	4	553	1735	2454
Kirklees	1	531	1712	2530
	3	527	1679	2475
	4	517	1747	2475

Wakefield	1	556	1801	2570
	3	553	1796	2553
	4	551	1836	2550
All 3 Boroughs	1	550	1753	2538
	3	546	1734	2498
	4	541	1772	2492

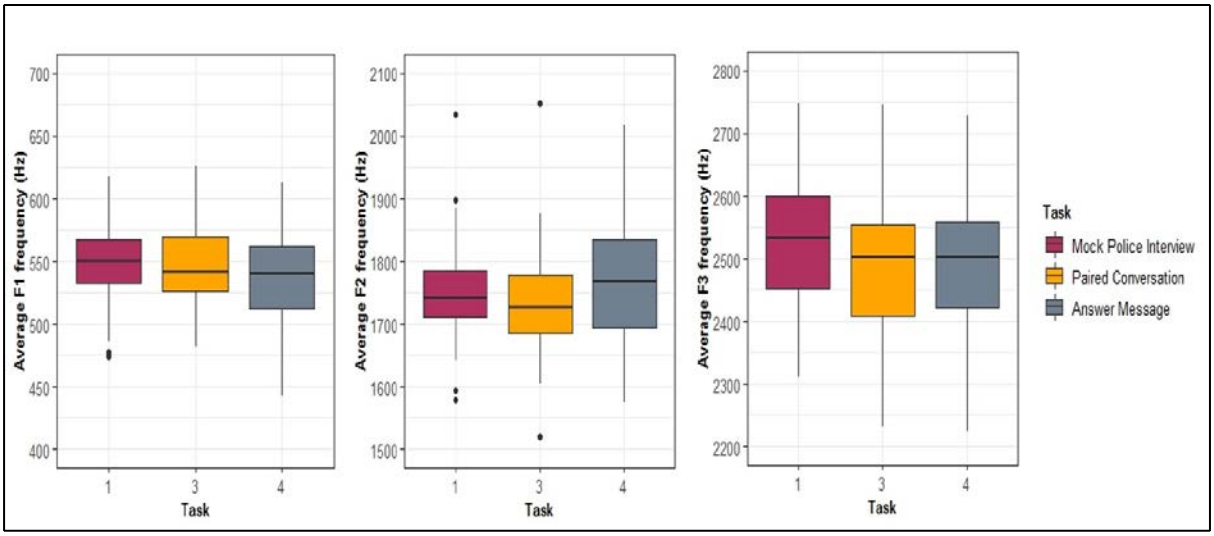


Figure 5.6. Average formant data across the three WYRED tasks.

Overall, it would appear that average F1 and F3 values are highest in Task 1 and lowest in Task 4, however, the difference in raw formant values is relatively small. In terms of F2, the average values are highest in Task 4 and lowest in Task 3. There is no obvious explanation for why this would be the case. It should be noted that there were high levels of between-speaker variation in terms of FACE formant values, and therefore the general trends that are observed in this section do not necessarily reflect the within-speaker differences across tasks for all speakers. Using the set of average midpoint F1~F3 values for each participant's FACE tokens across each of the three tasks, the smallest, largest, median and average change from the baseline to the two paired tasks were calculated. These values are presented in Table 5.5.

Table 5.5. Summary statistics relating to the change in average formant values across tasks.

		∆ in formant values (Hz)		
Statistic	**Task**	**F1**	**F2**	**F3**
Minimum	4 → 1	0.47	1.24	1.91
	4 → 3	0.29	1.04	2.12
Maximum	4 → 1	58.37	185.35	178.11
	4 → 3	61.07	221.69	170.91
Median	4 → 1	15.56	37.03	53.59
	4 → 3	15.90	54.21	23.26
Average	4 → 1	21.54	49.59	62.59
	4 → 3	18.90	65.96	44.72

Figure 5.7 presents the average BDM transformed F1 and F2 of each participant's FACE vowels across tasks, colour-coded based on the borough that they are from. The average F1 and F2 of the researcher's FACE vowels in each of the Task 1 mock police interviews are also presented. As each participant interacted with another participant during Task 3, it is not possible to distinguish the participant and interlocutor data in this task. In this figure, it can be seen that while the average FACE formants are in the same general area across all three tasks, there are some changes in terms of both F1 and F2 within vowels produced by participants from each of the three boroughs. This figure also illustrates that in Task 1 the researcher's FACE vowels are further back than those of the participants in this task.

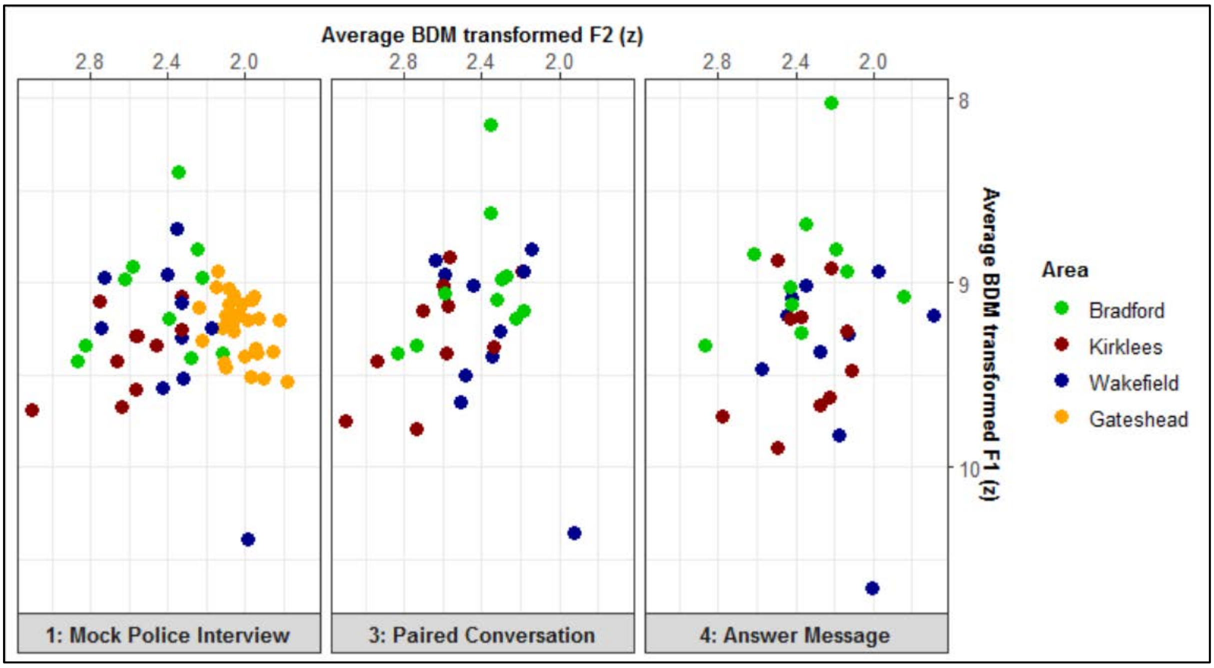


Figure 5.7. Changes in average FACE vowels across tasks.

5.5.2. How consistent are FACE productions within tasks by speaker?

An analysis of FACE F1~F3 SDs revealed that the participants' FACE productions varied in terms of how consistent they were across each task, with there being a general tendency for higher levels of variation in the two paired tasks (Tasks 1 and 3) than in the baseline task (Task 4), overall. Comparisons of SDs across tasks at the group level using linear mixed effects analyses showed that the SD of F2 was significantly influenced by the effect of speaking task. Taking Task 4 as the intercept in the models, TASK affected F2 SD ($\chi^2(2) = 41.27$, $p < 0.001$), increasing it by about 39 Hz ± 6 (standard errors) for Task 1 and increasing it by about 44 Hz ± 6 (standard errors) for Task 3. However, the results of the linear mixed effects analyses showed that TASK did not have a significant effect on F1 or F3 SD values. Additionally, BOROUGH was not found to be a significant predictor in the models for F1~F3 SD values. Full summary tables for the F1~F3 SD models are included in Appendix 3.

Table 5.6 provides a summary of the SDs of the FACE formant values across Tasks 1, 3 and 4, in each of the boroughs of Bradford, Kirklees and Wakefield, as well as the three areas combined. In terms of the overall SD values across areas, in the majority of cases the SDs are lowest in Task 4 compared to the two paired tasks. However, there are a few exceptions, for example Kirklees F1 SDs are essentially equal across all tasks, and the SDs for F3 are lowest in Task 1 for Bradford and when evaluating all three boroughs together.

Table 5.6. Standard deviations of FACE formant values across Tasks 1, 3 and 4, in the boroughs of Bradford, Kirklees, Wakefield and all three areas together.

Area	Task	F1 SD (Hz)	F2 SD (Hz)	F3 SD (Hz)
Bradford	1	53	148	141
	3	59	144	164
	4	49	130	177
Kirklees	1	47	155	174
	3	43	165	189
	4	44	139	168
Wakefield	1	62	173	173
	3	59	179	163
	4	54	127	150
All 3 Boroughs	1	56	163	165
	3	56	171	176
	4	52	139	170

The boxplot in Figure 5.8 visualises the SDs of the formant data for all participants across all three tasks. Please note that the y-axes for each of the F1~F3 subplots have been customised within a range the best suits the SD values and therefore they are not all the same. It can be seen that the SDs of FACE formant values vary across Tasks 1, 3 and 4. In terms of F1, the median SD is almost equal across the three tasks, the upper quartile is highest in Task 3 and the lower quartile is lowest in Task 4. These results show that generally speaking the F1 SDs are highest in Task 3 and lowest in Task 4, indicating that the FACE F1 values were most consistent during the baseline task and there may have been more speech accommodation taking place during Task 3 than Task 1. In terms of F2, it can be seen that SD values were distributed relatively similarly in the two paired tasks while the SD values were much lower in Task 4. Again, this suggests that FACE F2 values were most consistent during the baseline task. With regards to F3, the upper quartile of the SD values were similar across all three tasks but the lower quartile was much lower for Task 4 than for Tasks 1 and 3. The median SD value was highest for F3 during Task 3, indicating that the degree of speech accommodation in terms of F3 may have been highest during this task when considering all participants together.

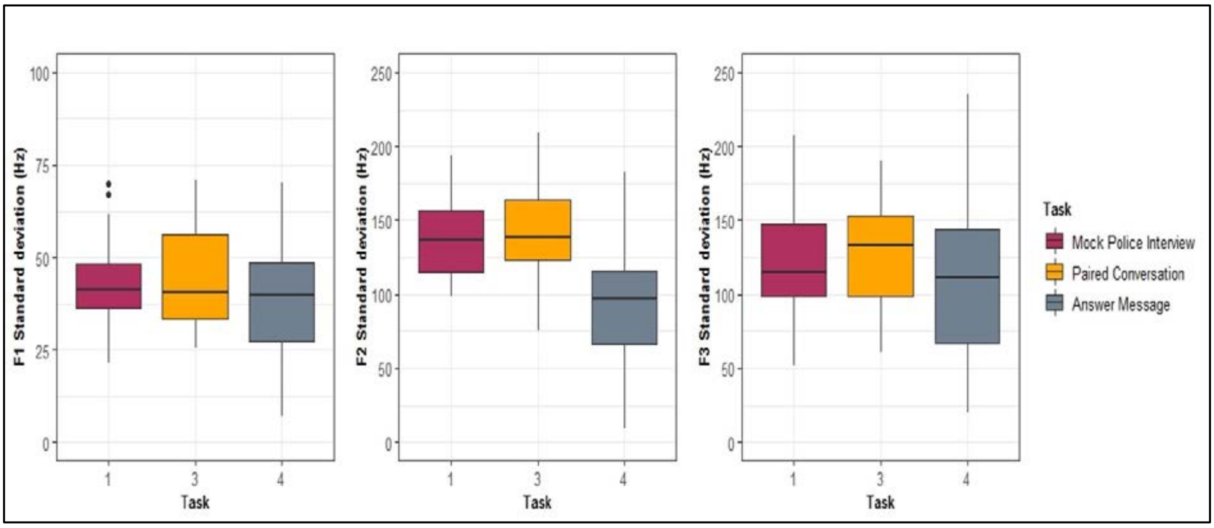


Figure 5.8. Standard deviations of formant data across the three WYRED tasks.

It can be seen in Figure 5.8 that the range of SD values across participants was greatest in Task 4, with some participants displaying a very low amount of variation across FACE tokens and some with very high SD values. This could in part be due to the fact that the number of FACE tokens produced during Task 4 was less consistent across participants than in Tasks 1 and 3, where most participants produced between 30-35 tokens.

Table 5.7 presents a summary of the number of participants who had higher SDs in either F1, F2 or F3, during the paired tasks than in the baseline task. As mentioned previously, in cases where there were higher levels of variation in the paired tasks than in the baseline task, this was taken to indicate that accommodation may have occurred during the paired task. There were eight participants who had higher SDs in all three formants during Tasks 1 and 3 than in Task 4 and there was just one participant (#006) who had higher SDs for F1~F3 in Task 4 than in the two paired tasks. Overall, it can be seen that out of the thirty participants, the majority displayed higher levels of variation in their FACE realisations in the two paired tasks than in the baseline task where no interlocutor was present. This suggests that perhaps some form of speech accommodation took place for these speakers over the course of the paired tasks;

whether that be convergence towards their interlocutor or divergence away from their interlocutor. However, it must be reiterated that the SDs in Task 4 were based on a much lower number of FACE tokens than in the two paired tasks, and therefore these values may be more heavily by influenced outliers and/or by the effects of varying phonetic environments.

Table 5.7. Number of participants, out of 30, with higher SDs in Tasks 1 and 3 than Task 4.

Formant	Task 1 > Task 4	Task 3 > Task 4
F1	19 (63%)	21 (70%)
F2	26 (87%)	25 (83%)
F3	26 (87%)	25 (83%)

In order to investigate how the participants' FACE productions relate to their interlocutor's productions, the following section presents findings in relation to the Euclidean distances between interlocutors at various points in time throughout the investigation.

5.5.3. How do the participants' FACE productions change as a result of exposure to their interlocutor?

In order to investigate how the participants' FACE productions may have been influenced by their interlocutor's productions, a set of *difference in distance* (DID) values were obtained for each participant, across the two paired tasks. As described in Section 5.4.3, the measures of phonetic accommodation represented by DID values were calculated by subtracting the *original distance* (average *baseline distance)* from each of the *paired distances*. Negative DID values represent convergence towards the interlocutor during the paired task and positive DID values represent divergence away from the interlocutor during the paired task. A DID value of 0 indicates no change as a result of auditory exposure to the interlocutor.

Overall, participants tended to diverge away from their respective interlocutors more than they converged towards them and levels of divergence appeared to be stronger during the mock police interview task than the casual paired conversations. However, high levels of between-speaker variability were observed with regards to the participants' accommodation behaviour. In the below sub-sections, the group trends in accommodation in FACE productions are first presented by comparing DID values between paired tasks and then secondly by comparing DID values from the first half to the second half of each paired task. In Section 5.5.3.2, the accommodation behaviour of a selection of participants are presented in order to illustrate the high levels of between-speaker variability.

5.5.3.1 Group results

5.5.3.1.1 Accommodation behaviour across paired tasks

Table 5.8 provides summary statistics in relation to all 30 participants' baseline distances, paired distances, DID values, and the percentage of tokens which had negative DID values (indicating a reduction in distance between interlocutor pairs). These statistics are presented separately for Task 1 and Task 3.

Table 5.8. Summary accommodation statistics for all participants, across Task 1 and Task 3.

	Statistic	Baseline distance (z)	Paired distance (z)	DID (z)	% tokens with DID < 0
	Average	0.70	0.84	0.09	46%
Task 1	**SD**	0.42	0.46	0.47	19%
Interlocutor	**Range**	0.05 - 2.22	0.04 - 3.57	-1.21 - +2.93	11% - 75%
	N	268	967	967	30
	Average	0.77	0.84	0.05	47%
Task 3	**SD**	0.41	0.51	0.46	15%
Interlocutor	**Range**	0.05 - 2.12	0.02 - 3.26	-1.71 - +2.52	18% - 71%
	N	268	881	881	30

The summary presented here suggests that overall participants diverged away from their interlocutor more than they converged towards them, during both of the paired tasks. However, when taking into account the relative baseline and paired distance figures, the average DID values are considered to be relatively small. The figures in the final column of Table 5.8 demonstrate that all participants converged on some of their FACE tokens during both of the paired tasks; however, based on the average percentage of tokens converged, it would appear that participants typically diverged away from their interlocutor on more tokens than they converged.

Figures 5.9 and 5.10 present histograms with overlaid density plots of each participants' averaged DID value across the two paired tasks, respectively. The dashed vertical line in each figure denotes the 0 DID point of no change. Both graphs show that the average DID values are relatively mixed, with participants falling on the positive and negative side of the scale across both of the paired tasks. In Task 1, 11 out of the 30 participants had average DID values below 0, and 14 participants converged in at least 50% of their FACE tokens. In Task 3, 12 participants had average DID values below 0, and 13 participants converged in at least 50% of their FACE tokens. Based on the statistics presented in Table 5.8 and the visualisations in Figures 5.9 and 5.10 it would appear that there is a large amount of between-speaker variation with regards to the accommodation behaviour displayed across the paired tasks. While some participants do not deviate far from the point of no change, the average DID values for others reflect more extreme changes from the baseline to the paired tasks.

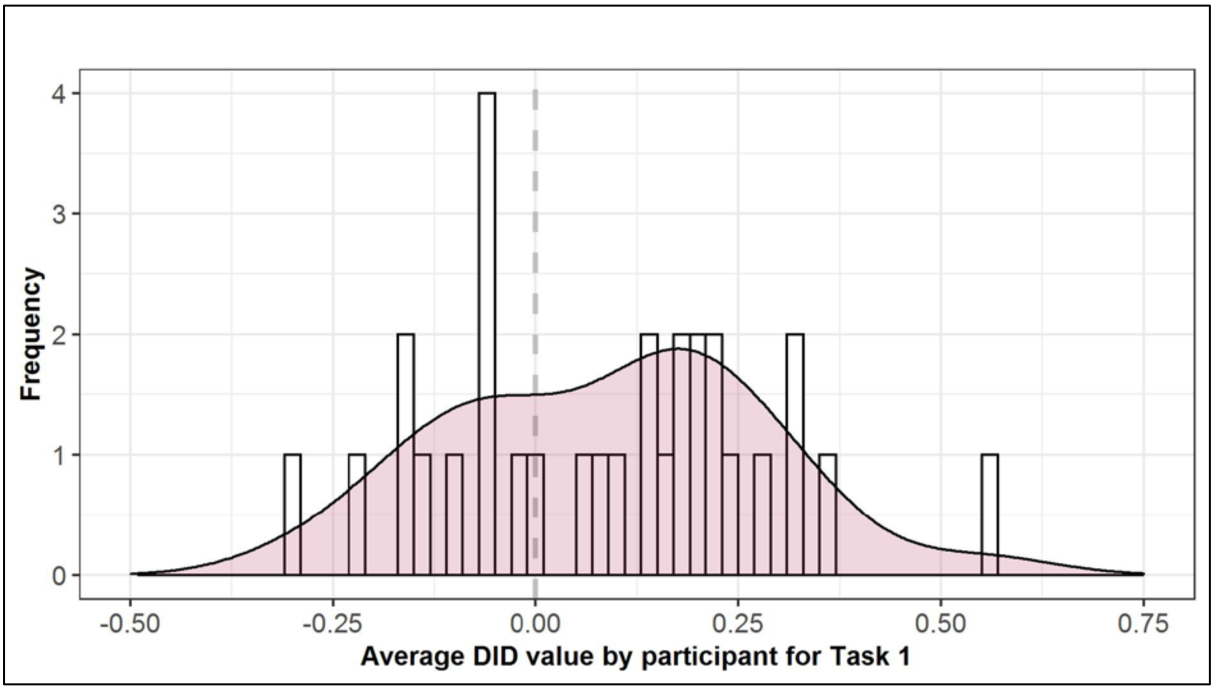


Figure 5.9. Histogram of each participant's averaged DID value in Task 1.

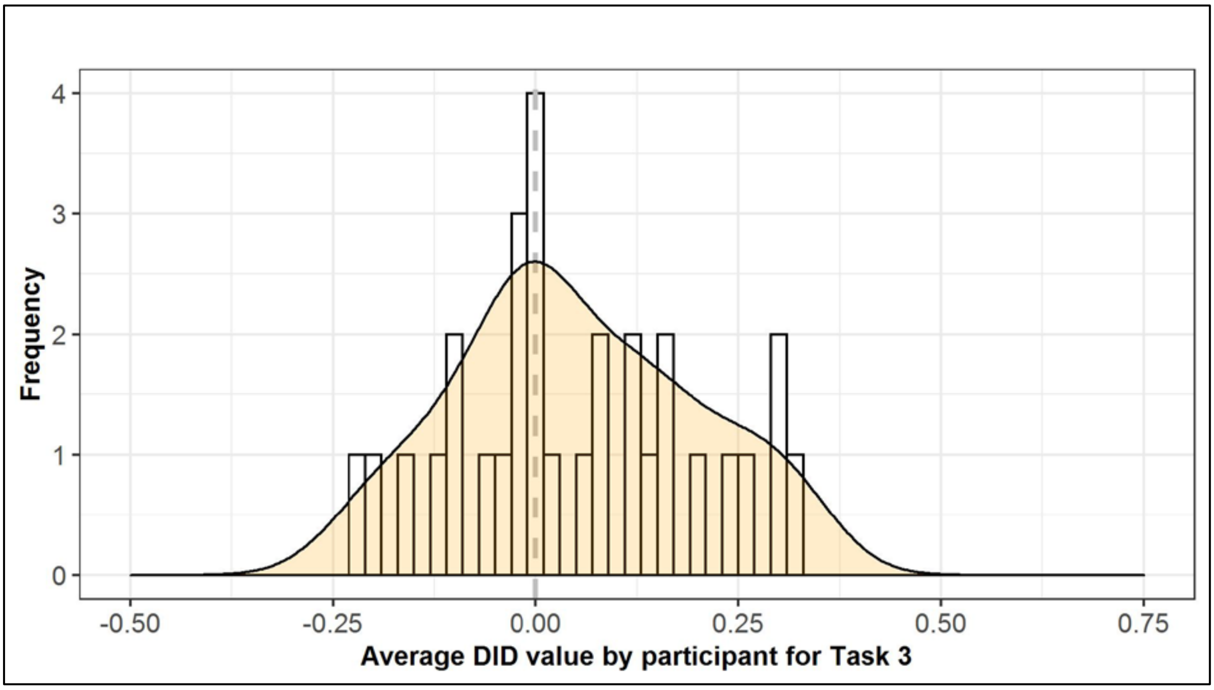


Figure 5.10. Histogram of each participant's averaged DID value in Task 3.

In order to determine whether there were overall effects of accommodation in each of the paired tasks across all participants, two Wilcoxon signed rank tests were conducted to determine whether DID values were significantly above 0. A Bonferroni correction adjusted the significant alpha level for the two comparisons to $p=0.025$. In Task 1, the DID values were significantly above 0 (Median = 0.05, SD = 0.47; $V = 261844$, $p < 0.001$). In Task 3, the DID values were not significantly above 0 (Median = 0.02, SD = 0.46; $V = 198587$, $p = 0.055$). These results indicate that there was a trend towards divergence in both paired tasks; however, the effect of divergence was only significant in Task 1. These findings are in line with what can be seen in Figures 5.9 and 5.10; while the peak in the Task 1 density plot is around 0.25 indicating divergence, the peak in the Task 3 density plot is around 0 suggesting maintenance.

In order to determine whether or not the participants' average DID values in Task 1 were correlated with their average DID values in Task 3, a Pearson's product-moment correlation coefficient was computed. The results show that the average DID values across the two tasks were positively correlated to a significant degree ($r(28) = 0.62$, $p < 0.005$). This suggests that participants who diverge in one task are more likely to diverge in the other than not, and similarly participants who converge in one task are more likely to converge in the other.

Comparing across the two paired tasks it can be seen that the average DID values were generally slightly higher in Task 1, meaning that levels of divergence were greater during this task. However, as many of the participants displayed high levels of variation in their DID values across the tokens in each of the two paired tasks, it is possible that each participant's average values do not adequately represent the differences between the two tasks. For instance, if a participant were to have converged strongly on some FACE productions and diverged heavily on others, the average value would cancel out the positive and negative values, and falsely indicate that the participant neither converged nor diverged. It is also possible that accommodation behaviour may have varied slightly across participants from each of the three separate boroughs and therefore by grouping all participants together when comparing across tasks, some patterns may be obscured.

Table 5.9 presents a summary of the average DID values and percentage of tokens in which the participants converged during Tasks 1 and 3, across each of the three West Yorkshire boroughs and the region as a whole. Here, the percentage of convergence is based on the proportion of individual tokens with DID values below 0. Overall, it would appear that Bradford participants tended to converge more than Kirklees and Wakefield participants, in Task 1 and Task 3. Kirklees participants converged the least across both tasks, based on the average DID values and the percentage of tokens with negative DID values.

Table 5.9. Summary of the average DID values and % of tokens in which phonetic convergence occurred, for each borough and all three areas together, across Task 1 and Task 3.

	Task 1		**Task 3**	
Area	**Average DID (z)**	**% Convergence**	**Average DID (z)**	**% Convergence**
Bradford	-0.01	55%	-0.01	54%
Kirklees	0.19	38%	0.13	37%
Wakefield	0.11	44%	0.04	49%
All 3 boroughs	0.09	46%	0.05	47%

In order to compare the DID values between tasks, taking into account any regional differences in the results and effects of phonetic environment, a linear mixed effects analysis was performed using all of the individual DID values. A description of the model used for this analysis is provided in Section 5.4.3.2. The results showed that there were no significant differences across tasks in terms of DID values. With regards to the control variables included in this model, it was determined that BOROUGH did not have a significant influence on the DID values, although the results were approaching significance ($\chi^2(2) = 5.664$, $p = 0.059$). However, the DID values were significantly affected by PHONETIC ENVIRONMENT: ($\chi^2(7) = 117.35$, $p < 0.001$). A full summary table for the DID value model is included in Appendix 3.

The violin plots in Figure 5.11 illustrate the variation in DID values from tokens across the different phonetic environments included in this study. Overall, it can be seen that the DID values for the main reference group (group 0) range from approximately -0.5 to +0.4 with a median value just below 0. This means that in the majority of FACE tokens there was a mix of convergence towards and divergence away from the participants' interlocutors. Groups 2 and 6, corresponding to FACE tokens produced before a nasal and in the lexical items *make* and *take*, are very similar to the main reference group. Groups 3 and 5, corresponding to FACE tokens produced after a glide and a liquid, were significantly higher than the main reference group meaning that participants tended to diverge away from their interlocutor to a stronger degree in these contexts. In contrast to this, participants tended to converge to a stronger degree in FACE tokens produced after a liquid (group 4) compared to the main group. It can also be seen that the distribution of DID values are similar across groups 1, 3 and 7, with a wide range in DID values from -0.75 to +1.25. It is important to note that although the variation across phonetic environments of FACE can be taken into account for the participants, this cannot be accounted for in the interlocutors' speech because their FACE productions are represented by an average of the F1 x F2 values from all of their FACE productions.

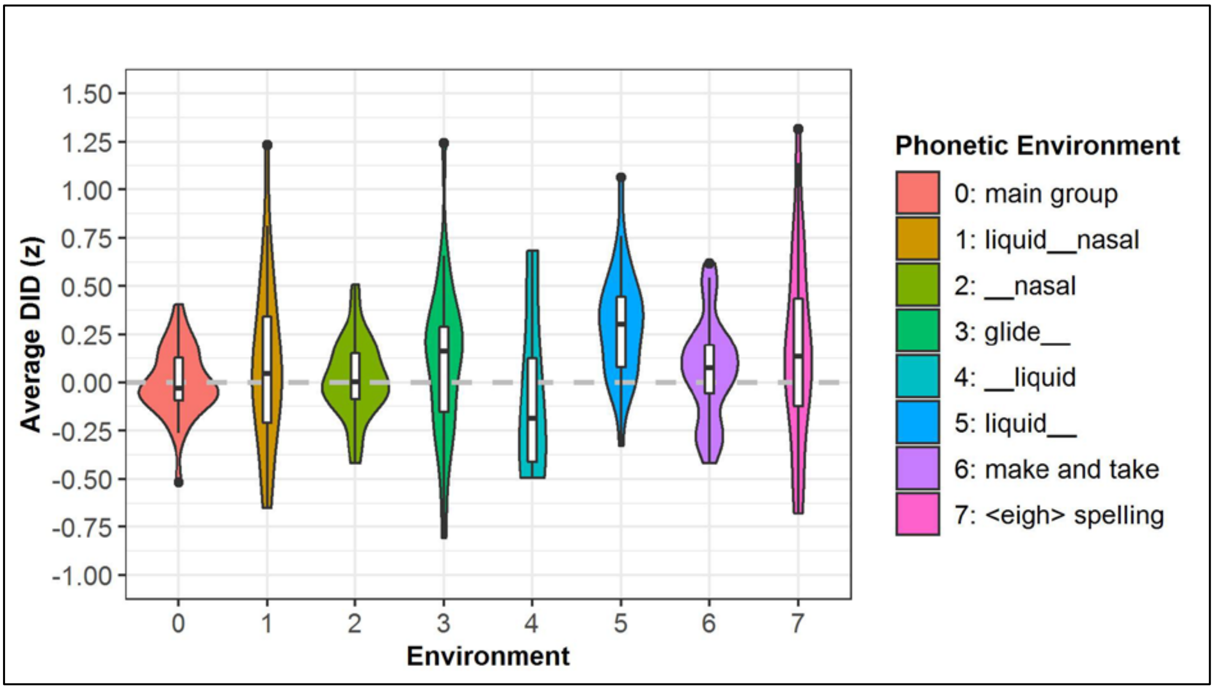


Figure 5.11. DID values across phonetic environments.

5.5.3.1.2 Accommodation behaviour over the course of a paired task

The final linear mixed effects analyses, conducted to examine the effects of exposure to the interlocutor over time on DID values, revealed that there were no significant differences across the two LATENCY conditions. This was the case in Task 1 and Task 3. This finding indicates that participants did not become significantly more or less similar to their interlocutor following more exposure. The DID values in Task 1 were significantly influenced by the control variable of BOROUGH. Taking Bradford as the intercept in the model, BOROUGH had a significant effect ($\chi^2(2) = 6.401$, $p < 0.05$), with DID values increasing by 0.21 ± 0.08 (standard errors) for Kirklees and increasing by 0.12 ± 0.08 (standard errors) for Wakefield. However, BOROUGH did not have a significant effect on the DID values in Task 3. The DID values in Task 1 were significantly affected by PHONETIC ENVIRONMENT ($\chi^2(7) = 79.903$, $p < 0.001$), as were the DID values in Task 3 ($\chi^2(7) = 41.346$, $p < 0.001$). Overall, these results indicate that changes in distance between interlocutor pairs over the course of the interaction can be better explained by the phonetic context of the FACE token rather than the extended exposure to the interlocutor. Full summary tables for the Task 1 and 3 DID models are included in Appendix 3. While it is not possible to discuss the results of all 30 participants in detail, the results of a selection of individuals are presented in the following sub-section to illustrate the range of different ways in which participants adapted their speech across the three tasks.

5.5.3.2 Selection of individual results

The results in the previous section showed that there were fairly high levels of between-speaker variation in terms of both average DID values and percentage of tokens in which participants converged. While some participants converged towards their interlocutor more than they diverged in either one or both of their paired tasks, others tended to diverge to a higher degree than they converged. One participant who displayed relatively high levels of convergence in both Task 1 and Task 3 was participant #006. In Task 1, 74% of the FACE tokens that this participant produced were a smaller distance from the interlocutor's average FACE realisation than the participant's original distance from the interlocutor during the baseline task, and their average DID value was -0.15 in Task 1. In Task 3, participant #006 converged in 71% of FACE tokens and had an average DID value of -0.22. Figure 5.12 presents participant #006's average transformed FACE vowel in the baseline and paired tasks, as well as their interlocutors' average FACE vowel in each of the paired tasks. The starting point of the two arrows represents the vowel in the baseline task (average transformed F1 and F2), while the end point represents the vowel in the paired task (average transformed F1 and F2). The maroon arrow represents the difference between the baseline and Task 1, while the yellow arrow represents the difference between the baseline and Task 3. The maroon circle labelled "Researcher" represents the Task 1 interlocutor's average transformed FACE vowel during the mock police interview with participant #006 and the yellow square labelled "#012" represents the average transformed FACE vowel of participant #006's interlocutor during the casual paired conversation, participant #012. Note that the x- and y-axes are reversed in line with standard conventions, in order to better represent the vowel space.

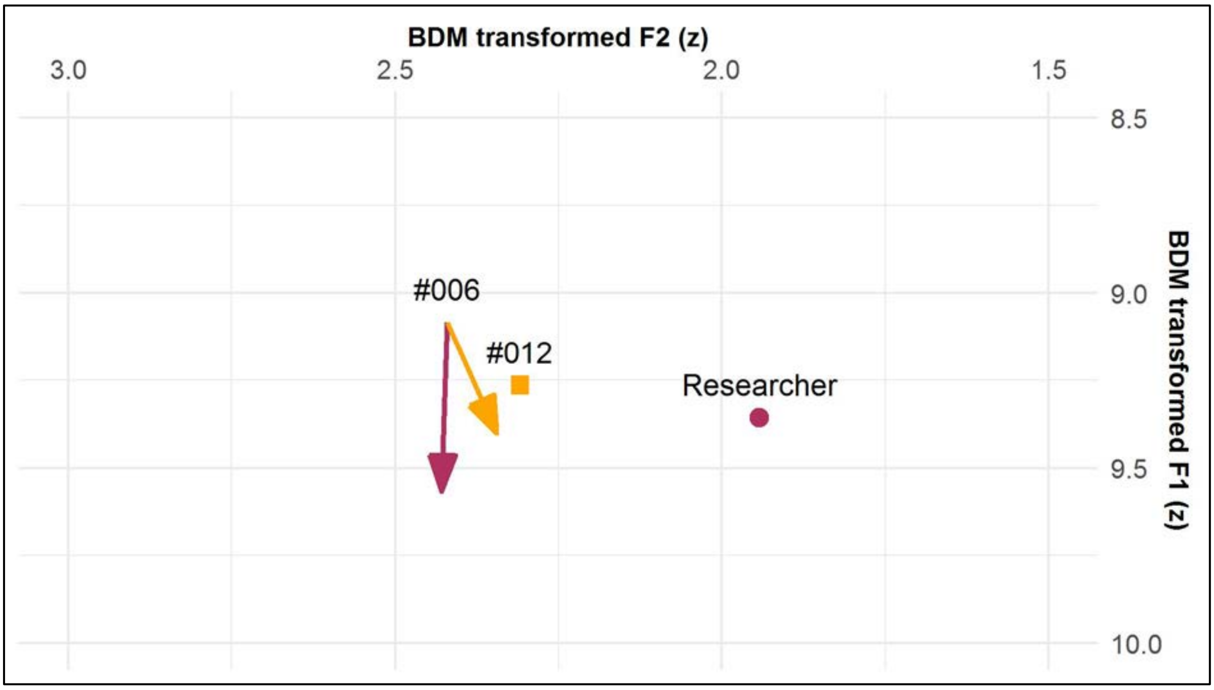


Figure 5.12. BDM transformed plot of #006's average FACE vowel across tasks compared to their interlocutors' productions.

In Figure 5.12 the change from participant #006's average FACE realisation in the baseline task to the average realisation in the two paired tasks can be observed. Firstly, if we focus on the maroon arrow, representing the change from the baseline to Task 1, it can be seen that F1 increases to become closer to the Task 1 interlocutor and F2 increases very slightly to become further away from the interlocutor. The overall distance between interlocutors is slightly reduced in Task 1 and therefore it can be considered that participant #006 converged towards the interlocutor during this task. If we now focus on the yellow arrow, representing the change from the baseline to Task 3, it can be seen that F1 increases and F2 decreases to become closer to the Task 3 interlocutor in both dimensions. It would again appear that the overall distance between interlocutors is reduced during this paired task and therefore participant #006 can be said to have converged in this task. Comparing the two paired tasks, the original distance between participant #006 and the Task 1 interlocutor was larger than that of participant #006 and the Task 3 interlocutor and this pattern was maintained even after convergence had taken place during the paired tasks.

If we compare the corresponding data for participant #006's partner in Figure 5.13, we can see that participant #012's accommodation behaviour can broadly be categorised as 'maintenance'. The format of this graph is the same as that in Figure 5.12, except that the maroon circle labelled "Researcher" represents the Task 1 interlocutor's average transformed FACE vowel during the mock police interview with participant #012 and the yellow square labelled "#006" represents the average transformed FACE vowel of participant #048's Task 3 interlocutor during the casual paired conversation. In terms of participant #012's average transformed F1 and F2 values, there is fairly little difference from the baseline task to Task 1, although there is more of a change in Task 3 particularly in F1. Participant #012 converged in 49% of FACE tokens in both Task 1 and Task 3 and had an average DID value of 0.14 and 0.07, respectively. Although positive DID values would suggest divergence overall, the magnitude of the difference from the baseline to Task 1 is small. Furthermore, while the average transformed formant values shown in Figure 5.13 indicate that participant #012 is converging towards his interlocutor, the positive average DID value indicates divergence. Accordingly, it can be concluded that this participant remained fairly consistent in both tasks overall.

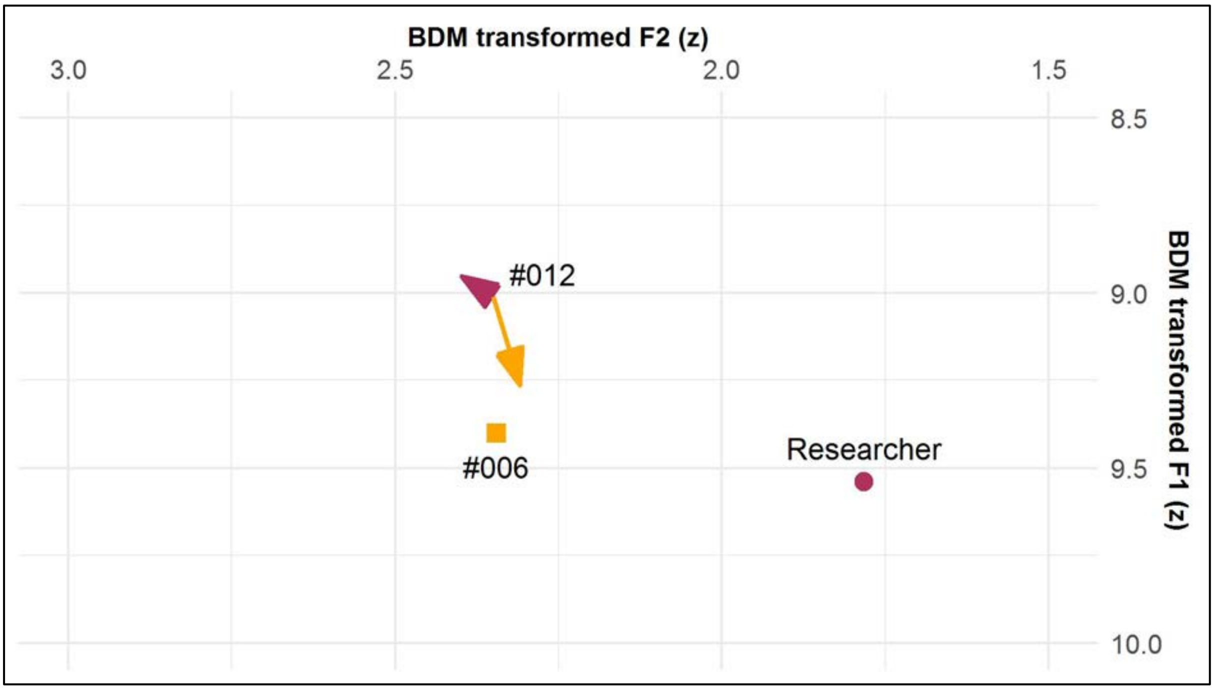


Figure 5.13. BDM transformed plot of #012's average FACE vowel across tasks compared to their interlocutors' productions.

In contrast to participants #006 and #012, participant #048 displayed relatively high levels of divergence in both Task 1 and Task 3. In Task 1, only 21% of the FACE tokens that this participant produced were a smaller distance from the interlocutor's FACE realisation than the participant's original distance from the interlocutor during the baseline task, and their average DID value was 0.55 in Task 1. Similarly in Task 3, participant #048 only converged in 29% of FACE tokens and had an average DID value of 0.25. Figure 5.14 presents participant #048's average transformed FACE vowel in the baseline and paired tasks, as well as their interlocutor's average FACE vowel in the paired tasks. In this graph, the maroon circle labelled "Researcher" represents the Task 1 interlocutor's average transformed FACE vowel during the mock police interview with participant #048 and the yellow square labelled "#036" represents the average transformed FACE vowel of participant #048's interlocutor during the casual paired conversation. Please note that the range on the x-axis has been customised within a range the best suits the transformed F2 values and therefore it is not that same as that in Figures 5.12 and 5.13.

Figure 5.14. BDM transformed plot of #048's average FACE vowel across tasks compared to their interlocutors' productions.

In Figure 5.14 the change from participant #048's average FACE realisation in the baseline task to the average realisation in the two paired tasks can be observed. Looking at both the yellow and maroon arrows it can be seen that F1 decreases to become closer to the interlocutors in Task 1 and Task 3 but F2 increases strongly to become further away from the interlocutors. The overall distance between interlocutors in Task 1 and in Task 3 is greatly increased and therefore it can be considered that participant #048 diverged away from the interlocutors during their respective tasks. Comparing the two paired tasks, the original distance between participant #048 and the Task 1 interlocutor was smaller than that of participant #048 and the Task 3 interlocutor and this pattern was maintained after divergence had taken place during the paired tasks. It is interesting to note that this participant's average transformed FACE production was very similar in Task 1 and Task 3, despite being far away from the two respective interlocutors.

As has been previously mentioned, when examining speech accommodation in spontaneous, conversational speech it must be acknowledged that it is possible that the speech of the model talker (or in this case the interlocutors) may be influenced by the participant's speech in addition to the participant being influenced by their interlocutor. If we consider the findings for participant #036 (the partner of #048 in Task 3) we can see that their average transformed F1 and F2 values were also different to their baseline FACE realisations during Task 3, suggesting that both participants in this conversation accommodated. Figure 5.15 presents participant #036's average transformed FACE vowel in the baseline and paired tasks, as well as their interlocutor's average FACE vowel in the paired tasks.

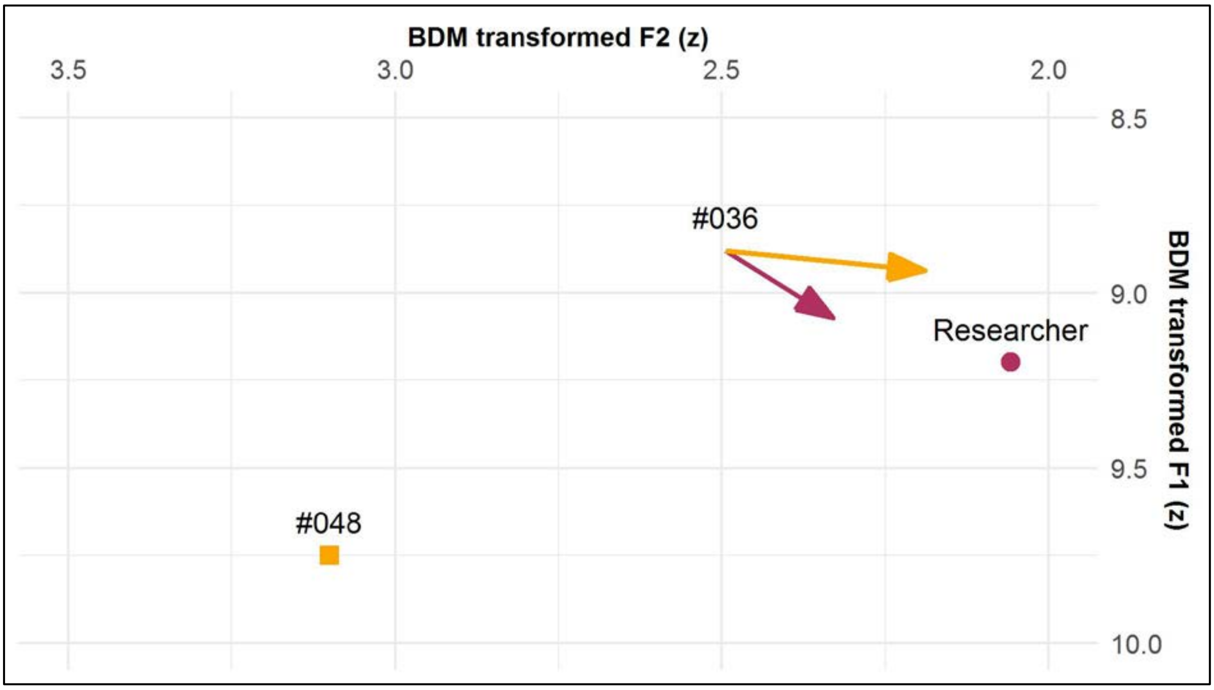


Figure 5.15. BDM transformed plot of #036's average FACE vowel across tasks compared to their interlocutors' productions.

In Figure 5.15 the change from participant #036's average FACE realisation in the baseline task to the average realisation in the two paired tasks can be observed. Firstly, if we focus on the maroon arrow, representing the change from the baseline to Task 1, it can be seen that F1 increases and F2 decreases, to become closer to the Task 1 interlocutor in both dimensions. The overall distance between interlocutors is reduced in Task 1 and therefore it can be considered that participant #036 converged towards the interlocutor during this task. If we now focus on the yellow arrow, representing the change from the baseline to Task 3, it can be seen that F1 increases slightly to become closer to the Task 3 interlocutor (participant #048) but F2 decreases to become much further away from the interlocutor. In this case, the overall distance between interlocutors is increased during Task 3 and therefore participant #036 can be said to have diverged in this task. Comparing the two paired tasks, the original distance between participant #036 and the Task 1 interlocutor was smaller than that of participant #036 and the Task 3 interlocutor. This pattern was emphasised during the paired tasks, as participant #036 converged towards the Task 1 interlocutor whereas both participants diverged away from one another during Task 3.

With reference to Figures 5.14 and 5.15, it can be seen that the average transformed FACE values of participants #036 and #048 during their respective baseline tasks were closer to one another than they were during Task 3. Both participants diverged away from each other during their paired task, with the average transformed F2 values increasing for participant #048 and decreasing for participant #036. The Task 1 interlocutor's average transformed FACE production is relatively consistent across the four interviews exemplified in Figures 5.12-5.15. This indicates that the researcher may have been less influenced by the speech of the participants during the mock police interviews than the participants were during the casual paired conversation.

In order to evaluate the findings of these four particular participants in more depth, an auditory analysis of each of their FACE tokens across all three tasks was undertaken. During this analysis, six different FACE variants were identified which varied according to vowel quality and length. These can be seen in Figure 5.16, in which the distribution plots for these participants and their respective interlocutors across tasks are presented.

Firstly, it is worth noting that the Task 1 interlocutor (the researcher) largely realised FACE as diphthongal variants [eɪ] and [ɛɪ], and this was consistent across all four of the participant's interviews. This can be seen in the second column of each of the plots shown in Figure 5.16. The researcher tended to use a higher proportion of diphthongal FACE variants than the West Yorkshire participants, however, she did use the monophthongal variants [eː] and [ɛː] occasionally but she never used the short variant [ɛ]. There were also a number of instances where FACE tokens were realised with a wider diphthong [æɪ]. With regards to the participants, it can be seen that they all used a similar range of variants but the proportional distribution of variants differed across participants and tasks.

Figure 5.16. FACE variant distribution plots for four participants and their respective interlocutors across tasks.

Participant #006 had relatively similar distributions of FACE variants across the three tasks, with the majority of FACE tokens being monophthongal. The main change from the baseline Task 4 to the two paired tasks was that this participant produced some [ɛɪ] tokens when talking to interlocutors who had [ɛɪ] as their most common FACE variant, despite this variant not being present at all in Task 4. It can also be seen that when talking to an interlocutor who does not use the short [ɛ] variant, participant #006 uses proportionally fewer [ɛ] tokens in Task 1 than in the other two tasks. The Task 3 partner of participant #006, participant #012, mainly realised FACE as [ɛɪ] across all three tasks. Participant #012's auditory analysis results are broadly in line with the acoustic findings that he had fairly consistent average F1 and F2 values across tasks. The slight trend reported above towards divergence in Task 1 could be related to the increased usage of [ɛ] in this task, as the Task 1 interlocutor did not use this variant. Relatedly, the trend towards convergence in Task 3 may be linked to the increased usage of [ɛː] and [eː], as these variants were both used by the participant's partner in this task.

In Figure 5.16, it can be seen that participant #036 mainly tends to produce monophthongal variants of FACE. In Task 4, all tokens were realised as [ɛː] whereas there was much more variability in the two paired tasks. One reason for this could be that there was more opportunity for variation as more FACE tokens were produced over a longer period of time in each of the paired tasks. Interestingly, while the acoustic analysis suggested that participant #036 converged towards their interlocutor in Task 1 and diverged away from their interlocutor in Task 3, the opposite trends would appear to be true based solely on the auditory analysis results. While participant #036's FACE distributions appear very different to his Task 1 interlocutor, his distributions are relatively similar to those of his Task 3 interlocutor, participant #048.

Participant #048 mostly realised FACE as [eɪ] in Task 4, however, when interacting with his Task 1 interlocutor (who has high usage of [eɪ]) his proportional use of this variant dramatically reduced. Additionally, he had a higher proportion of monophthongal tokens during Task 1, in opposition to his interlocutor. It should be noted that lexical effects and the influence of the phonetic context were not controlled and accounted for in this auditory analysis and therefore it is likely that some variation was a result of these factors. For instance, it was noted that half of the [ɛ] tokens produced by these participants occurred in the words *make* and *take* (15 out of 31 tokens).

In this section evidence has been presented of participants converging towards their interlocutor in terms of their FACE productions across both paired tasks (participant #006), diverging away from their interlocutors in both tasks (participant #048), maintaining their original distance (participant #012) and also converging strongly in one task while diverging strongly in another (participant #036). Appendix 4 provides further summary formant plots for all 30 participants which display their average transformed FACE productions across tasks with their interlocutors' FACE productions marked for reference.

5.6. Discussion

This section discusses the results of the research questions addressed in this investigation in more detail, and briefly summarises how the findings of this investigation relate to those of previous accommodation studies. The implications of the findings for FSC casework and researchers in the fields of forensic speech science and sociophonetics more broadly are also explained.

One of the primary motivations for examining the West Yorkshire FACE vowel in this thesis was that it was believed to be socially salient in this region and have high levels of variation across the West Yorkshire boroughs. In Chapter 4, it was observed that local level variation was present in West Yorkshire specifically in terms of the front/back dimension of the FACE vowel. The investigation presented in this chapter has examined the extent to which 30 speakers from West Yorkshire accommodate in their productions of the FACE vowel. Based on the findings of previous studies which had investigated phonetic accommodation in vowel formants and other variables which were considered to be socially salient, it was anticipated that some form of accommodation would occur in FACE vowel productions. It was also predicted that there would be differences in terms of the levels of accommodation present across the two paired tasks. However, no specific prediction was made in terms of the expected direction of accommodation in each of the tasks or with regards to which task the participants would converge towards their interlocutors the most. The reason for this was that previous studies had found conflicting findings about when convergence was more likely to occur and, due to the semi-spontaneous nature of the data used in this study, there were too many variables to consider to be able to make a reliable prediction. Unsurprisingly, the general results of the present investigation reveal that the effects of speech accommodation are highly speaker specific and consequently cannot be straightforwardly predicted based on the speaking task alone.

5.6.1. Summary of findings

This investigation has defined accommodation to refer to changes that occur across multiple speaking situations involving different interlocutors. Although it is not possible to (completely) separate the effects of speaking style from the influence of the interlocutor, the findings of this study indicate that the FACE vowel does appear to be influenced by accommodation, for many speakers. This sub-section provides a summary of the answers to each of the three research questions (RQs) addressed in this investigation.

RQ1: What is the influence of the task on the FACE vowel productions?

An analysis of the formant values of all FACE tokens produced by the 30 participants across the three tasks revealed that there were significant acoustic differences across tasks in terms of F2 midpoint values, but not in terms of F1 or F3. This result suggests that in the case of FACE, the F2 dimension relating to vowel front/backness may be most likely to be influenced by phonetic accommodation. This finding may be in part influenced by the fact that FACE F2 values were found to be most regionally marked across this group of speakers, with significant differences in F2 being observed across participants from the boroughs of Bradford, Kirklees and Wakefield. Although it seems unlikely that participants were aware that their FACE vowel front/backness was altering across different tasks, it may be the case that shifts occurred subconsciously in this particular dimension due to the variation present more generally in West Yorkshire.

Overall, the average amount by which participants adapted their FACE realisations across tasks in terms of raw hertz was typically relatively small. The largest differences were in terms of

F2, which increased on average by approximately 60 Hz from the baseline task to the two paired tasks. However, it is possible that these changes may still be perceivable to the listener, as Flanagan (1955) reported that changes in F1 or F2 of as little as 3% could be detected by listeners in a vowel discrimination task involving synthesized vowels. Given that the average F2 for this group of speakers was approximately 1750 Hz, a change of 60 Hz (3.4%) would lie close to the threshold for just-noticeable differences. In order to comprehensively assess this for each of the individual participants in this study, a perceptual analysis would be required to complement the acoustic findings. However, this was beyond the scope of this study.

RQ2: How consistent are FACE productions within tasks by speaker?

It was found that for the majority of participants, FACE productions were most consistent during the baseline answer message task (Task 4) where there was no interlocutor present. For these participants, higher levels of variation in FACE formant values were observed in the two paired tasks, which could be taken to indicate that the participants' FACE realisations were altering over the course of these tasks as a result of the influence of the interlocutor. However, it was acknowledged that the lower SDs in Task 4 may have been a by-product of there being a comparatively small number of FACE tokens available in this task and consequently less variety in terms of phonetic context of the FACE vowels.

All of the linear mixed effects analyses that included the phonetic environment as a fixed effect showed that this factor was a significant predictor in the models, suggesting that the FACE vowel is highly affected by the surrounding phonetic context. For instance, in Chapter 4 it was found that both F2 and F3 values of FACE tokens that occurred following a liquid (e.g. *play*, *Rachel*) or following a glide (e.g. *way*, *Wakefield*) were significantly lower than the reference level in the models. The average F2 and F3 values for the reference level in the models were 1777 Hz and 2558 Hz, respectively. Whereas, the equivalent values in FACE tokens following a liquid were 1637 Hz and 2436 Hz, respectively, and those for FACE tokens following a glide were 1676 Hz and 2394 Hz. These findings were not considered to be surprising as it is well recognised that liquids often cause lowering of F2 for front vowels (Ladefoged & Maddieson, 1996; Ladefoged, 2001). However, it was also noted that in these particular phonetic contexts, the DID values were generally higher, meaning that participants appeared to diverge from their interlocutor to a greater extent than in most other environments. This finding suggests that accommodation behaviour may be to some extent conditioned by the phonetic context of FACE. Although this variable was taken into account when analysing F1~F3 and DID values, the imbalance of tokens from each phonetic environment across tasks could not be accounted for when comparing the standard deviations of formant values.

RQ3: How do the participants' FACE productions change as a result of exposure to their interlocutor?

The effects of exposure to the interlocutor on FACE productions were considered by comparing baseline productions with those from the two paired tasks, as well as by comparing participant productions over the course of each paired task. Across the mock police interviews and the casual paired conversations a general trend was observed whereby participants tended to diverge away from their interlocutor more than they converged, with this only happening to a significant extent during the police interview task. This finding is perhaps surprising as we might have predicted that convergence would be more likely to occur just by virtue of the fact that during the paired tasks the pairs of interlocutors took part in the same task and therefore used similar speaking styles; whereas in the baseline task participants left an answer message in a time-pressured situation. Nevertheless, it should be recognised that there were high levels of between-speaker variability in terms of how the participants accommodated during the two paired tasks, and some participants did not fit the general trend as they converged more than they diverged.

An evaluation of the DID values revealed that levels of divergence were slightly higher in Task 1 than Task 3 overall; however, a linear mixed effects analysis showed that the DID values were not significantly different across tasks. When evaluating whether the DID values were significantly different from zero, in order to assess whether convergence or divergence could be deemed to have occurred overall, it was found that the effect of divergence was only significant in the mock police interview task. This meant that, in comparison to a casual conversation with someone similar to themselves in many respects, participants tended to diverge to a stronger extent when they were interacting in a more formal situation, involving asymmetric power dynamics, where they were being interviewed by a community outsider of the opposite sex.

Despite there being high levels of between-speaker variability in terms of accommodation, leading to overall differences across tasks being relatively marginal, it is still worth considering why levels of divergence might have been higher in Task 1 than in Task 3. Previous research has indicated that phonetic alignment or convergence is typically considered to be a technique for rapport building and has been found to occur more often during positive interactions (Byrne, 1971; Lee et al., 2010). As the mock police interview task required the participants to lie in order to avoid incriminating themselves, it is plausible that this may have caused participants to react negatively towards the interviewer. As a consequence of this, participants may have diverged away from the police interviewer as a means to accentuate individual differences or express distain (Bourhis & Giles, 1977; Shepard et al., 2001). This would explain the higher levels of divergence during Task 1. In contrast to this, Task 3 involved a casual paired conversation where it was in the participants' interest to maintain a positive interaction as they were aware that they had to hold a conversation with their partner for twenty minutes. For this reason, it is not surprising that levels of divergence were lower in Task 3 than in Task 1, as the participants had more motivation to cooperate with their interlocutor; however, it is still surprising that participants did not converge to a greater degree overall. Although there was a slight trend towards divergence in this task, the peak of the density plot shown in Figure 5.10 was only slightly above 0, suggesting that participants were more likely to produce FACE in a way that was similar to in their baseline task, as opposed to converging or diverging.

Another explanation for why the overall effect of accommodation may not have been as strong in the casual conversations as it were in the mock police interview recordings, could be that participants were generally less consistent in how they behaved as a group due to the fact that they were all interacting with different interlocutors. It may be the case that some of the participants evoked convergence more than others in the paired conversations. Equally, a higher degree of consistency in the interviews could be a consequence of all participants being interviewed by the same female researcher from Gateshead.

In all cases, during the Task 3 conversations participants were paired with male speakers from the same West Yorkshire borough as themselves and were of a similar age. Pardo et al. (2018b) and Pickering & Garrod (2013) suggested that speakers who are more similar to one another to begin with are more likely to converge. Pardo et al. also stated that based on this assumption, "same-sex pairs of talkers should converge more than mixed-sex pairs." (2018b, p. 4). This might help to account for why higher levels of divergence were found in Task 1 than in Task 3, as Task 1 involved mixed-sex pairs of speakers from different locations. With regards to the FACE vowel specifically, the West Yorkshire participants were expected to have relatively similar productions (largely in the vicinity of [eː] or [ɛː]) whereas the researcher from Gateshead had a wider range of variants, with predominant use of the diphthongal [eɪ] and [ɛɪ] variants and open-mid monophthongal variants in the vicinity of [ɛː] and [eː], depending on the phonetic context of the token.

Regarding the effect of latency, participants did not seem to converge or diverge to a stronger degree as a result of more exposure to the interlocutor over the course of the paired tasks. A possible explanation for this could be that the two paired tasks were not long enough for any effects of exposure time to become apparent. Perhaps if the interlocutor pairs had spent more time together, they may have become more familiar with one another which could have led to increased convergence. Alternatively, the lack of latency effects might be related to the fact that there was a general trend towards divergence, overall. It could be the case that divergent behaviour is generally more immediate than convergence. However, I am not aware of any previous studies that have tested this claim empirically.

When comparing the DID values between speakers, it is evident that speech accommodation behaviour is highly speaker dependent. This is demonstrated by the high levels of between-speaker variability across tasks both in terms of the degree and direction of accommodation. It was considered whether some of the variability between participants could have been related to which borough they were from. The summary of the average DID values and percentage of phonetic convergence presented earlier in Table 5.9 showed that there were some differences in results when the participants were grouped according to borough. Furthermore, the participants' DID values in Task 1 were shown to be significantly different across the three boroughs of West Yorkshire, with the average DID values for Bradford participants displaying a tendency to converge while the Kirklees and Wakefield participants tended to diverge. However, this pattern was not observed in the DID data from Task 3 or when considering the DID values from both paired tasks together. For this reason, it seems unlikely that local level regional differences play a strong role in determining accommodation behaviour.

As the participants included in this study were similar in terms of age, gender and educational background, it is possible that accommodation is conditioned by more speaker specific influences. A number of previous studies have explored the relationship between accommodation behaviour and a wide range of social factors and have shown that convergence can sometimes be linked to particular personality traits. For instance, listener traits such as openness, conscientiousness, attention-switching and rejection sensitivity have been put forward as factors which can influence the amount of convergence in an interaction (Aguilar et al., 2016; Yu et al., 2013). It has also been proposed that convergence rates can be influenced by social biases about how a participant feels about their interlocutor (Babel, 2010) and how attractive a speaker finds the model talker (Byrne, 1971). It is possible that internal factors such as these may have contributed to why certain participants were found to converge to varying degrees towards their interlocutor, while others tended to diverge. Unfortunately, examining the relationship between accommodation and personality traits/opinions is outside the scope of this study, and not possible as the WYRED project did not elicit this information.

5.6.2. Implications

5.6.2.1 Forensic speech science

The results of this study have shown that generally speaking the FACE vowel productions were more consistent in the speaking task where no interlocutor was present. It may therefore be useful for forensic experts to pay attention to whether or not speech samples that are to be analysed in FSC casework involve a single speaker or multiple speakers. In cases where the speaker of interest is interacting with other people, higher levels of within-speaker variation may be observed in features such as vowel realisations which could potentially be accounted for by the effects of accommodation. This may be of particular importance in FSC cases involving a mismatch between samples whereby one sample involves an interactive conversation whereas the other sample involves monologue-style speech that we might expect to find in an answer message or a video recording made by a single individual.

Additionally, the findings of this study indicate that more extreme mismatches across samples (in terms of interlocutors and speaking styles) may result in higher levels of accommodation. In a real FSC case, being interviewed by a police officer may evoke a phonetic response in some suspects resulting in their speech being realised in a different way than usual. Moreover, if a suspect was interviewed by an officer with a different accent to their own this could potentially lead to changes in their speech that would not be observed if they were to be interviewed by an officer with a similar accent. However, one of the main take-home messages from the findings presented in this chapter is that speech accommodation is highly speaker specific.

Results suggest that some speakers are more heavily influenced by the speech of their interlocutor and the effects of differing speaking tasks than others. It would seem that, for the majority of speakers, accommodation in vowel formants is unlikely to present a major issue for FSC casework. Although within-speaker variation in FACE productions is evident both within and between the different tasks, for most participants this variation would not be substantial enough to lead an expert to interpret the findings as providing evidence in support for the view that different speakers had produced the FACE tokens. For this reason, the effects of accommodation on vowel formants that occur across speaking situations involving different interlocutors and speaking styles would appear unlikely to have a significant impact on a FSC conclusion. However, there are some speakers who accommodate more strongly and, as exemplified via auditory analysis in Section 5.5.3.2, within-speaker variability exists for some speakers in terms of FACE variants used across different tasks. Furthermore, if differences are observed within one parameter it is conceivable that differences in other parameters may co-occur. Consequently, it is possible that accommodation in multiple parameters could affect an expert's overall assessment of similarity and it remains to be seen how other phonetic parameters are influenced by accommodation within the participants included in this study.

The findings of this investigation have also highlighted a more general point, slightly removed from the effects of accommodation, which is the need to account for the phonetic environment in which the feature under examination is observed. The results of the linear mixed effects analyses presented in this chapter and in Chapter 4 revealed that the phonetic environment of FACE had a significant effect on both the midpoint formant values of the vowel as well as the DID values relating to accommodation behaviour with respect to this feature. For this reason, it would be useful to control the phonetic environment when comparing vowel tokens across samples, wherever feasible. However, it is acknowledged that this is not always possible when dealing with evidential samples as they are often limited in terms of the speech available.

5.6.2.2 Sociophonetics

This study has also added to the phonetic accommodation literature by applying established acoustic-phonetic techniques to measure speech accommodation in semi-spontaneous speech data. The findings of this study are in line with many other accommodation studies which have reported relatively subtle effects of phonetic convergence with high variability across speakers (Pardo et al., 2018b). In this study, the effects of different speaking styles across tasks cannot be straightforwardly unpicked from the influence of the interlocutors and therefore it is not explicitly clear which of these factors had the greatest effect on the levels of within-speaker variability in FACE vowel productions. Overall, it appears that the F2 dimension of the West Yorkshire FACE vowel is most heavily affected by the mismatches across tasks in terms of style and interlocutor and this may be linked to the fact that F2 has been found to be most regionally marked across West Yorkshire in Chapter 4. This finding potentially supports the proposition put forward by Trudgill that accommodation is more likely to occur in speech parameters with high levels of awareness associated with them (1986, p. 11).

From a methodological perspective, it is recommended that sociolinguistic/phonetic researchers take both speaking task and the influence of the interlocutor into account when dealing with spontaneous speech data. When researching speech accommodation specifically, it is also suggested that, as accommodation behaviour seems to be highly speaker specific, information relating to the participants' personality and attitudes should be collected using methods such as the administration of the Implicit Association Task (Greenwald, Mcghee, & Schwartz, 1998) and attractiveness ratings (as applied in Babel, 2009, 2010). More generally when conducting sociophonetic research, it is important that researchers are mindful of the potential influence that their own speech may have on their participants during sociolinguistic interviews and that steps are taken to maintain consistency across participants, wherever possible. Furthermore, it is important to acknowledge the possibility that participants recorded as part of the same interview may not be entirely independent of one another.

5.7. Conclusion

This chapter has built on the findings of the previous chapter (which investigated how FACE was produced across the boroughs of Bradford, Kirklees and Wakefield) by exploring how the West Yorkshire FACE vowel is realised in different stylistic contexts. By considering within-speaker variability in FACE formant values across three separate tasks, it was possible to evaluate how susceptible this phonetic feature is to the influence of speech accommodation.

The findings of this study have demonstrated that the way in which the West Yorkshire participants realise the FACE vowel can be affected by the speaking task and the influence of the interlocutor. Participants tended to diverge away from their interlocutors during both of the paired tasks to a greater extent than they converged, overall. However, there were participants who converged towards their interlocutors, and some were fairly consistent in their FACE productions across tasks. Based on the levels of accommodation observed in this study, it seems unlikely that accommodation in vowel formants would have a substantial impact on a FSC conclusion for most speakers. Nevertheless, it is clear that accommodation is something that experts need to be aware of when carrying out a FSC analysis, as some speakers are more heavily influenced by the effects of speech accommodation than others. The following chapters will consider how word-medial, intervocalic /t/ is influenced by speech accommodation and the effect of social salience will be taken into account to assess how this links to accommodation behaviour.

6. Variation in /t/ across West Yorkshire

6.1. Introduction

This chapter presents an investigation into how word-medial, intervocalic /t/ is realised across West Yorkshire[4]. One motivation for examining this parameter is that it has been found to vary throughout the UK as well as in West Yorkshire specifically, with glottal variants of /t/ becoming increasingly more common. This investigation explores which variants are currently in use across the West Yorkshire boroughs of Bradford, Kirklees and Wakefield. One aim of this investigation is to provide an up-to-date account of intervocalic /t/ usage for an area that has received relatively little attention in recent years. Another aim is to evaluate to what degree local level variation exists across West Yorkshire with respect to this feature.

This chapter is divided into 7 sections. The first section provides a review of findings relating to how /t/ productions have been shown to vary across the UK, as well as how this feature can vary according to social and linguistic factors. In Section 6.2, the research questions and hypotheses are set out. In Section 6.3, a brief summary of the data that has been examined in this investigation is presented and Section 6.4 provides an explanation of the methods used in this study. Section 6.5 presents the results to each of the research questions posed in this chapter by setting out how /t/ is realised in West Yorkshire and the extent to which it varies across the three boroughs. In Section 6.6, the results of this investigation are discussed in detail and the chapter is concluded in Section 6.7.

6.1.1. Regional variation

The way in which /t/ is produced by speakers from different parts of the UK has been an area of great interest amongst sociolinguists and phoneticians for many years. This sub-section summarises the numerous different /t/ variants that have been reported across various parts of the UK and specifically in the area of West Yorkshire. Although there have been relatively few sociophonetic studies that have focussed on West Yorkshire English, Broadbent (2008),

[4] A preliminary version of the investigation presented in this chapter has been published in Earnshaw & Gold (2019).

Hughes, Trudgill, & Watt (2012) and Petyt (1985) have each presented accent descriptions for boroughs within West Yorkshire which have included details of how /t/ is realised.

Petyt (1985) examined the accents of Bradford, Halifax and Huddersfield and found that two non-standard forms of intervocalic /t/ were in use word-finally: the glottal stop [ʔ] and "a sort of 'linking r' (realised by either the usual frictionless continuant or a flap)" (Petyt, 1985, p. 151). This meant that the phrase *get out* might be realised as either [geɹaʊʔ], [geɾaʊʔ] or [geʔaʊʔ]. Petyt also observed the use of [ʔ] for word-medial intervocalic /t/, though it was least frequent and most highly stigmatised in this context. Based on his findings, Petyt concluded that although the glottal stop was considered a stigmatised variant, it appeared to be increasing in its frequency in all environments, with higher usage of [ʔ] word-finally than word-medially and evidence of [ʔ] replacing the older non-standard [ɹ]/[ɾ] variants. It was also suggested that speakers from the city of Bradford were leading the change, with usage of [ʔ] for /t/ diffusing from this area to the towns of Halifax and Huddersfield.

In Broadbent's examination of /t/ within speakers from Morley, South Leeds, usage of [ʔ] for /t/ appeared to be on the increase while use of [ɹ] for intervocalic /t/ (commonly referred to as *t*-to-*r*; Wells, 1982, p. 370) appeared to be declining. However, examples of [ɹ] were present both word-finally and word-medially to a lesser extent. It was also noted that in some instances, /t/ was realised as a tapped variant [ɾ], in line with the findings of Petyt for other parts of West Yorkshire. Broadbent described both T-glottaling and the process of *t*-to-*r* as forms of lenition, with T-glottaling being the dominant lenition form in West Yorkshire (Broadbent, 2008, p. 164). More recently, T-glottaling has been reported in Bradford in word-final contexts (Hughes et al., 2012, p. 106). However, in this description of Bradford English there was no mention of /t/ being realised as either [ɹ] or [ɾ] and no description of word-medial /t/ was provided. Based on the accounts relating specifically to West Yorkshire, it would seem that word-medially we might expect to find [ʔ], [ɹ] and [ɾ] in addition to standard forms of /t/, with the glottal stop being the most prevalent non-standard variant. Word-finally, [ʔ] is anticipated to be the most common variant for /t/.

In recent years, the replacement of the voiceless alveolar plosive [t] for [ʔ] in non-initial position has been extensively studied by phoneticians and sociolinguists. Despite traditionally being described as one of the two most heavily stigmatised features of British English pronunciation alongside H-dropping (Milroy, Milroy, & Hartley, 1994, p. 4), it would appear that these negative connotations are broadly declining. A recent study by Alderton (2020) found that T-glottaling is highly salient in listener perceptions and can elicit strong reactions based on the variable's social associations (2020, p. 43). However, it was suggested that social meanings associated with T-glottaling "cannot be reduced to merely a vague notion of 'stigma'" (2020, p. 40) as they can vary depending on a speaker's social and linguistic characteristics, and in particular their gender (2020, p. 45). Over the last century, T-glottaling has become prevalent across most parts of the UK, at least in some contexts, and its usage is still thought to be on the rise. According to Watson, "the presence of the glottal stop as a realisation of /t/ is arguably one of the most common phonological processes in that it occurs in many varieties of British English" (2006, p. 56). Furthermore, Smith & Holmes-Elliott state that "glottal replacement is largely considered to be a 'torchbearer' of geographical diffusion" (2017, p. 1).

At this point, it is necessary to draw an important distinction between the term "*T-glottaling"* and "*glottalisation"*. In this investigation, *T-glottaling* is used to refer to what some have also termed *glottal replacement*; i.e. "the variable realisation of underlying /t/ with an auditorily distinct glottal stop [ʔ], which either masks the oral plosive release or replaces the plosive entirely" (Smith & Holmes-Elliott, 2017, p. 1). This is in contrast to *glottalisation,* which is used in this study to refer to instances where there is an auditory impression of a double articulation involving adduction of the vocal folds, resulting in either a glottal closure or a period of creaky voice, prior to an oral closure. This type of articulation is also sometimes referred to as *glottal reinforcement* (cf. Docherty & Foulkes (1999, p. 54). In Docherty & Foulkes' acoustic analysis of a sample of 549 tokens from the Newcastle corpus, which on auditory analysis appeared to contain a glottal, they classified tokens into three categories: 'pure' glottal variants, and two subtly different types of glottalisation (1999, p. 57). In the present study, examples of /t/ being realised as [ʔ] are referred to as *glottal* whereas examples of [ʔ͡t] are termed *glottalised*. Table 6.1 presents a summary of areas across the UK where T-glottaling and glottalisation has been documented. Although this list is by no means exhaustive, it provides in insight into how widespread this non-standard feature has become.

Table 6.1. Regional distribution of glottal and glottalised /t/ variants.

/t/ variant	Region	Reference
[ʔ]	Belfast	McCafferty, 1999, p. 249
	Blackburn	Turton, 2017
	Bradford	Hughes et al., 2012, p. 106
	Buckie, Northeast Scotland	Smith & Holmes-Elliott, 2017
	Cardiff	Mees & Collins, 1999, p. 192
	Derby	Docherty & Foulkes, 1999, p. 50
	Edinburgh	Chirrey, 1999, p. 226; Schleef, 2013, p. 208; Hughes et al., 2012, p. 130
	Glasgow	Stuart-Smith, 1999a, p. 208 Stuart-Smith, 1999b, p. 188
	Hull	Williams & Kerswill, 1999, p. 147 Hughes et al., 2012, p. 109
	Lancashire	Hughes et al., 2012, p. 150
	Liverpool	Clark & Watson, 2016, pp. 40–41
	London English (South East & West)	Tollfree, 1999, p. 171; Schleef, 2013, p. 208 Hughes et al., 2012, p. 6
	Manchester	Baranowski & Turton, 2015, p. 305 Hughes et al., 2012, p. 116
	Middlesbrough	Hughes et al., 2012, p. 120
	Milton Keynes	Williams & Kerswill, 1999, p. 147
	Newcastle	Watt & Milroy, 1999, pp. 29–30 Docherty & Foulkes, 1999, p. 69
	Northumberland	Hughes et al., 2012, p. 155
	Norwich	Trudgill, 1999, p. 132
	Reading	Williams & Kerswill, 1999, p. 147
	Sandwell	Mathisen, 1999, p. 110
	Sheffield	Stoddart et al., 1999, p. 75
	West Yorkshire (Bradford, Halifax & Huddersfield)	Petyt, 1985, p. 149

[ʔ͡t]	Belfast	McCafferty, 1999, p. 249
	Edinburgh	Chirrey, 1999, p. 226
	Middlesbrough	Hughes et al., 2012, p. 120
	Newcastle	Watt & Milroy, 1999, pp. 29–30 Hughes et al., 2012, p. 68
	Norwich	Trudgill, 1999, p. 132
	Sandwell	Mathisen, 1999, p. 110
	Sheffield	Stoddart et al., 1999, pp. 75–76

In contrast to the ever-increasing use of T-glottaling and glottalisation of /t/, the realisation of /t/ as [ɹ] or [ɾ] appears to be more restricted in terms of areas where this feature is found. Wells (1982) introduced the concept of the "*t*-to-*r* rule" to refer to instances where /t/ is realised as a rhotic segment - often in the environment of a preceding short vowel. Wells described this as a "widespread but stigmatized connected-speech process in the middle and far north [of England], [involving] the use of /r/ instead of /t/ in phrases such as *shut up* [ʃʊɹˈʊp], *get off* [ˈgɛɹˈɒf]" (1982, p. 370). It was also noted that very occasionally the rule also applied word-internally, as in *what's the matter?* [ˈmaɹə] (1982, p. 370). Table 6.2 presents a summary of the areas where application of the *t*-to-*r* rule has been reported. In some areas, /t/ is replaced by [ɾ], which has been referred to by some scholars as *tapping* (Harris & Kaye, 1990; Wells, 1982) and by others as *flapping* (Hughes et al., 2012; Turton, 2017). In other areas, /t/ has been replaced by the approximant [ɹ]. It can also been seen that in areas such as West Yorkshire, both [ɹ] and [ɾ] have been reported. By comparing Table 6.1 and Table 6.2, it is also possible to see that in many areas both T-glottaling and application of the *t*-to-*r* rule are reported. These variants often change depending on the phonetic environment.

Table 6.2. Regional distribution of *t*-to-*r*.

/t/ variant	Region	Reference
[ɾ]	Belfast	McCafferty, 1999, p. 249
	Blackburn	Turton, 2017
	Glasgow	Stuart-Smith, 1999b, p. 188
	Hull	Hughes et al., 2012, p. 109

[ɾ]	Lancashire	Hughes et al., 2012, p. 150 Turton, 2017
	Leeds	Broadbent, 2008, pp. 145–146
	Leicester	Hughes et al., 2012, p. 102
	Liverpool	Clark & Watson, 2011, p. 530 Hughes et al., 2012, p. 113
	Sandwell	Mathisen, 1999, p. 110
	South East London English & South East London Regional Standard	Tollfree, 1999, p. 171
	West Yorkshire (Bradford, Halifax & Huddersfield)	Petyt, 1985, p. 151
[ɹ]	Derby	Docherty & Foulkes, 1999, p. 51
	Leeds	Broadbent, 2008, pp. 145–146
	Newcastle	Watt & Milroy, 1999, pp. 29–30; Docherty, Foulkes, Milroy, Milroy, & Walshaw, 1997, p. 292
	Sheffield	Stoddart et al., 1999, p. 76
	West Yorkshire (Bradford, Halifax & Huddersfield)	Petyt, 1985, p. 151

In addition to glottal and rhotic realisations of /t/, a wide range of other variants have been reported across the UK. Table 6.3 presents the distribution of some of the other /t/ variants that are currently in use, most of which can be grouped together on the basis that, like T-glottaling and tapping, they also involve lenition. Lenition is described by Clark & Watson as "a cover term given to a set of phonological processes that, among other things, turn phonological plosives into affricates, oral fricatives, glottal fricatives, or glottal stops" (2016, p. 35). For instance, affricated realisations of /t/ such as [t^{s}] and realisations involving an extended period of oral friction following the stop gap [tθ], have become heavily associated with Liverpool English, as have fully spirantised realisations of /t/ such as [s] and [h]. However, as can be seen in Table 6.3, some of these lenited productions of /t/ have also been reported in other parts of the UK including within some parts of Yorkshire. For this reason, it is possible that examples of this variant may also be present in West Yorkshire English.

Table 6.3. Regional distribution of further /t/ variants.

/t/ variant	Region	Reference
[tˢ]	Hull	Hughes et al., 2012, p. 109
	Liverpool	Hughes et al., 2012, p. 113
	Middlesbrough	Hughes et al., 2012, p. 120
	Sheffield	Stoddart et al., 1999, p. 76
	South East London Regional Standard	Tollfree, 1999, p. 170
	West Wirral	Newbrook, 1999, p. 97
[s]	Middlesbrough	Hughes et al., 2012, p. 120
	Liverpool	Hughes et al., 2012, p. 113
[tθ]	Liverpool	Clark & Watson, 2016, pp. 40–41
[θ]	Liverpool	Clark & Watson, 2016, pp. 40–41
[h]	Liverpool	Watson, 2006, p. 60 Clark & Watson, 2016, p. 40
[t̬]	Cardiff	Mees & Collins, 1999, p. 192
	Newcastle	Watt & Milroy, 1999, pp. 29–30
[k]	Manchester	Baranowski & Turton, 2015, p. 305

One particular variant that stands out in Table 6.3 is the realisation of /t/ as the velar plosive [k] in Manchester English. This has been reported to occur exclusively before syllabic /l/ in words such as *little* and *hospital*, although it is said to be highly stigmatised, increasingly rare and likely to fade in the next few generations (Baranowski & Turton, 2015, p. 305). The voiceless alveolar plosive [t] has not been included in Table 6.3, as this variant is believed to occur across almost all areas of the UK, in certain contexts. Tables 6.1-6.3 illustrate that there are a wide range of /t/ variants in use throughout the UK. It can also be seen that multiple variants are observed within individual areas. The patterns of use of each variant are therefore not only conditioned by regional variation but also by social factors as well as linguistic constraints. These effects are addressed in the following sub-sections.

6.1.2. Social variation

In their examination of /t/ in Derby and Newcastle, Docherty & Foulkes (1999) found that fine-grained phonetic features correlated with speakers' social characteristics, including age, gender and to some extent social class. For instance, in pre-vocalic position, glottal stops were more common in younger speakers in Derby and Newcastle, particularly within middle-class females (Docherty & Foulkes, 1999; Watt & Milroy, 1999). Whereas it was noted that older males in Newcastle largely used glottalised forms such as [ʔ͡t] (Docherty & Foulkes, 1999, p. 54). In the West Midlands, Mathisen (1999) also reported that the glottal stop was most common in younger speakers, and that females and middle-class speakers tended to use this variant most often. However, the same group of West Midlands speakers also used the glottalised variant [ʔ͡t] but to a lesser degree than [ʔ]. The tapped variant [ɾ] was reportedly the preserve of males, with frequencies of use increasing with age. Many other studies have also reported that T-glottaling is more common in younger speakers than older speakers (Baranowski & Turton, 2015; Smith & Holmes-Elliott, 2017; Stoddart et al., 1999).

In Glasgow, there was a higher overall usage of T-glottaling in younger speakers than older speakers, however, there was an interaction between age and social class meaning the pattern was more complex (Stuart-Smith, 1999b, p. 191). In contrast to some of the findings reported above, working-class speakers used T-glottaling more than middle-class speakers in both casual conversations and when reading a wordlist (Stuart-Smith, 1999b, p. 189). Baranowski & Turton (2015) also found that working-class speakers from Manchester were more likely to use [ʔ] than the middle-class speakers. Furthermore, the working-class males were the only speakers to show a considerable amount of T-glottaling in the most highly stigmatised context (intervocalic), suggesting that this social group may be leading the change towards T-glottaling in this context. In Reading, Hull and Milton Keynes, T-glottaling was said to be the norm among working-class people. Among middle-class young people, females used glottal replacement more in Reading and Hull, while in Milton Keynes the males had the higher frequency (Williams & Kerswill, 1999).

Stoddart et al. (1999) reported that [ʔ] was more frequently used in Sheffield by males than females whereas Stuart-Smith (1999b) found no quantitative gender differences in Glasgow.

Gavaldà reported that in SSBE T-glottaling was used “similarly by men and women, although in words such as *that is, get up, what if* it is used slightly more often by female speakers, since male speakers tend to prefer t-tapping in this context" (2016, p. 64). Smith & Holmes-Elliott (2017) found that older males used [ʔ] more often than older females; however, these gender differences were not present within the younger speakers who tended to use the glottal variant approximately 90% of the time.

The seemingly conflicting findings across different studies regarding the interaction between gender and usage of [ʔ] have previously been accounted for through reference to the local status of the variant and the stage of change within the speech community of interest (Smith & Holmes-Elliott, 2017, p. 5). In Alderton's (2020) investigation of perceptions of T-glottaling within teenagers from Hampshire, results indicated that high rates of alveolar /t/ production indexed positive traits for girls (such as popularity), whereas similar rates of [t] for boys had negative associations (such as being too intellectual or geeky) (2020, p. 43). More generally, it has been suggested that males have led the change towards T-glottaling when it has been regarded as the stigmatised form, whereas females have tended to adopt the innovative feature in cases where it is associated with supralocal norms. It has also been argued that gender differences reduce as rates of T-glottaling increase within the speech community (Smith & Holmes-Elliott, 2017, p. 5). In the present study, the aforementioned external factors are largely held constant in order to focus on how location and phonetic environment affect /t/ realisations across West Yorkshire.

6.1.3. Variation across phonetic environments

When analysing variation in /t/ for the purposes of comparing realisations between speakers, and groups of speakers, it is necessary to control for the phonetic environment in which this variable occurs. Previous studies of T-glottaling have found that a number of different linguistic factors condition both where [ʔ] can be produced and how frequently it occurs, including the position of /t/ within the word, the sound immediately preceding and following /t/, as well as the stress and prominence of the syllable in which it occurs. A selection of the key findings in relation to each of these aspects are summarised below.

There appears to be a general consensus that T-glottaling typically occurs more often in word-final contexts than word-medial contexts (Baranowski & Turton, 2015; Petyt, 1985; Schleef, 2013; Stoddart et al., 1999; Stuart-Smith, 1999b; Tollfree, 1999). Stuart-Smith (1999b) observed that older working class Glasgow speakers style-shifting from casual to formal styles, only replaced glottal stops with [t] in intervocalic position, and maintained their categorical use of [ʔ] pre-pausally and pre-vocalically. This finding was interpreted as evidence of intervocalic T-glottaling being more socially salient, which was said to make sense given that this is generally considered to be the most stigmatised position (Stuart-Smith, 1999b, p. 195). Fabricius provides evidence in support of this claim from a perceptual analysis of T-glottaling in which speakers rated pre-vocalic examples of [ʔ] as being less acceptable than pre-pausal and pre-consonantal examples (2002: 132). Petyt (1985) also reported that the glottal variant was most highly stigmatised in the intervocalic context. In addition to T-glottaling depending largely on the position of /t/ within a word, the phonological context also appears to heavily influence the linguistic patterning of this variant.

Hughes et al. (2012, p. 67) have proposed a constraint hierarchy whereby [ʔ] is said to occur most to least frequently in the following contexts:

1. Word-final pre-consonantal (*that man*)
2. Before a syllabic nasal (*button*)
3. Word-final pre-vocalic (*that apple*)
4. Before syllabic [l̩] (*bottle*)
5. Word-internal pre-vocalic (*better*)

However, the precise nature of the constraint hierarchy for glottal replacement has been shown to vary somewhat between locations (cf. Smith & Holmes-Elliott, 2017; Stuart-Smith, 1999b) and even across different generations within the same speech community (Smith & Holmes-Elliott, 2017). For instance, while Stuart-Smith (1999b) found that in Glasgow rates of T-glottaling in word-medial, ambi-syllabic contexts (number 4 in the hierarchy shown above) were much lower than in word-final contexts, Smith & Holmes-Elliott (2017) reported the opposite pattern in Buckie. A number of other studies have reported particularly high rates of T-glottaling when /t/ occurs before syllabic /l/ (Mathisen, 1999, p. 116; Watt & Milroy, 1999, p. 30) and therefore some might argue that this context should be positioned further up in the hierarchy.

Another linguistic constraint conditioning T-glottaling is stress. Previous studies of T-glottaling have shown that this phenomenon is highly sensitive to prominence patterns, with T-glottaling generally being blocked in foot-initial onset position, both word-initially (e.g. *tiny*) and word-internally (e.g. *attend*) (Tollfree, 1999, p. 172). Further to this, a generalisation was proposed which stated that in most locations T-glottaling is only an option where “the stress on the syllable following /t/ is less than that borne by the preceding syllable” (Tollfree, 1999, p. 172). However, some speakers in an advanced stage of this phonological process do realise /t/ as [ʔ] in the stressed onset context, although this generally only happens in a very restricted set of so-called –ee/-oo environments such as *thirteen, canteen, cartoon*, *tattoo*, (Harris & Kaye, 1990, p. 271). Examples of areas where speakers have been found to produce glottal tokens in this position include London (Harris & Kaye, 1990), Buckie (Smith & Holmes-Elliott, 2017) and Manchester (Baranowski & Turton, 2015). In London and Buckie, the non-standard variant was said to be frequently allowed in this context, whereas in Manchester it was reported that this seemingly advanced stage of glottaling was rare.

The phonetic environment also appears to play in role in conditioning when application of the *t*-to-*r* rule can apply. When Wells (1982) initially wrote about this phenomenon, it was suggested that /t/ could be realised as a rhotic segment when word-final /t/ is preceded by a short vowel and followed by a vowel-initial word. It was also noted that occasionally this could also apply word-medially (Wells, 1982: 370). Docherty et al. suggested that sentence stress may also have an effect, whereby “weakening to [ɹ] seems more likely to occur when the main phrasal prominence is not located on the syllable where /t/ is the rhymal consonant” (Docherty et al., 1997, p. 292). More recently, it has been proposed that realisations of /t/ as [ɹ] only tend to occur in a restricted set of specific lexical items. For instance, in Newcastle, [ɹ] only occurs in a restricted set of common verbs (e.g. *get*, *let*, *put*) and non-lexical words (Watt & Milroy, 1999, p. 30). In Liverpool, it was also observed that /t/ is only realised as [ɹ] in certain lexical items whereas the tapped variant [ɾ] has much more relaxed lexical constraints on its distribution (Clark & Watson, 2011). Some scholars have proposed usage-based models to account for the *t*-to-*r* phenomenon. For example, based on data from West Yorkshire, Broadbent (2008) proposed that *t*-to-*r* is more likely to be found in high-frequency words and high-frequency phrases. Clark & Watson (2011) extended this proposal by suggesting that *t*-to-*r* is not constrained simply by word frequency or phrase frequency alone, but by a combination of both word and phrase frequency. Gavaldà also observed that in SSBE "T-tapping appears more often in grammatical and highly frequent words rather than in lexical and less frequent words, both word-internally and across word boundaries" (2016, p. 64).

As part of an investigation of /t/ in Liverpool and the neighbouring areas of St Helens and Skelmersdale, a summary of the complex linguistic constraints that condition the variability between realisations of /t/ as [t] or as a [h] in Liverpool English was presented. It was claimed that [h] seemed to be constrained by a range of linguistic factors. This included utterance position, with [h] occurring only in pre-pausal position; lexical frequency, with high-frequency words favouring [h]; vowel length (in monosyllabic words, [h] is more common following a short vowel than a long vowel); and vowel stress (in polysyllabic words, [h] is more common following an unstressed vowel) (Clark & Watson, 2016, p. 36).

The present study focuses on intervocalic /t/, which has been reported as either [t], a glottal variant [ʔ], or a tapped or continuant /r/ in parts of West Yorkshire, in this specific phonetic environment (Broadbent, 2008; Petyt, 1985, p. 151). A decision was taken to focus solely on /t/ tokens that occurred in a word-medial position, as preliminary examinations of a subset of the spontaneous speech data considered in this investigation indicated that there were insufficient instances of intervocalic /t/ occurring between word boundaries for a robust analysis. An analysis of word-final /t/ was also considered, however, early indications again highlighted that there were very few pre-pausal tokens and there were insufficient tokens to be able to carry out a thorough analysis controlling for the effects of the following phonetic context. It was also noted that the previous literature cited above indicated that T-glottaling occurs much more commonly word-finally and that it is considered to be much less stigmatised (and hence less socially salient) in this context than word-medial /t/. For these reasons, it was not anticipated that there would be any local level variability across the three boroughs, nor was it expected that there would be within-speaker variability between speaking tasks. An initial pilot analysis of word-final /t/ confirmed that this variable was almost categorically realised as [ʔ] in this position, across all three speaking tasks.

Additionally, word-initial /t/ was not considered in this analysis, as the only variants expected in this position were [t] or [tʰ]. Furthermore, previous literature relating to West Yorkshire had not suggested that variation was likely to occur in this context. In the present investigation, all instances of word-medial, intervocalic /t/ were examined, with tokens separated into different groups by phonetic environment based on previous findings cited within the literature.

6.2. Research questions and hypotheses

This study aims to identify which allophones of intervocalic /t/ are currently in use across West Yorkshire. It also seeks to explore the extent to which /t/ varies between the boroughs of Bradford, Kirklees and Wakefield in order to determine if a local level identity is indexed through accent on a more fine-grained level than general Yorkshire English. Consequently, the research questions being addressed in this investigation are as follows:

1. How is word-medial, intervocalic /t/ realised across the metropolitan boroughs of Bradford, Kirklees and Wakefield?
2. Are there differences in /t/ realisations across the three West Yorkshire boroughs in question and if so, how extreme are these differences?

Based on the previous findings of sociolinguistic research involving /t/ in West Yorkshire, it is anticipated that the realisations [t], [ʔ], [ɹ] and [ɾ] will be observed across the different phonetic environments of word-medial, intervocalic /t/. Glottal productions are expected to be more common than rhotic realisations, in line with the trend observed by Broadbent (2008) whereby increased usage of [ʔ] appeared to be leading to a decrease in rhotic segments. As the participants included in this study are all young males, rates of T-glottaling are expected to be relatively high. It is also predicted that rates of T-glottaling will be higher in unstressed positions than in the stressed onset context and before syllabic /l/ than before vowels, therefore care will be taken to distinguish tokens from each of these categories.

With regards to the second research question, while it is anticipated that T-glottaling will be present across all three boroughs, it is expected that there may be some variation between boroughs in the extent to which participants use the glottal variant. In 1985, Petyt found that rates of T-glottaling were higher in Bradford than in Halifax or Huddersfield (Kirklees). At the time, the differences between areas were believed to be due to Bradford acting as a centre of innovation from which a linguistic change was diffusing. As T-glottaling is now much more widespread throughout the UK, this may no longer be the case. However, it does appear that some regional nuances still exist across the boroughs of Bradford, Kirklees and Wakefield, as evidenced by the fact that the FACE vowel was found to vary across boroughs in the investigation presented in Chapter 4. Furthermore, anecdotal accounts of accent variation across West Yorkshire boroughs were provided by WYRED participants during the recording process.

6.3. Data

In line with the previous chapters presented in this thesis, data for this analysis is from 30 young, male speakers from West Yorkshire. Each participant took part in three tasks and all tokens of word-medial, intervocalic /t/ produced during these tasks were analysed auditorily. The three tasks involved a mock police interview scenario, a casual paired conversation and a short task where participants were asked to leave an answer message. The tasks were all designed to elicit semi-spontaneous speech in forensically-relevant speaking situations. Data from all three tasks was analysed for this investigation.

All participants were male, aged 19-29 at the time of the recordings taking place and were all university educated. This meant that the social factors outlined in Section 6.1.2 that can influence /t/ productions, did not need to be accounted for during this investigation. However, it must be acknowledged that there may be an interaction between social characteristics (i.e. age, gender and socio-economic background) and location, which have not been considered in this investigation. For this reason, caution must be taken in generalising the findings of this study to the wider West Yorkshire community as a whole. For further details regarding the participants and the experimental design, please see Chapter 3.

6.4. Methodology

This section outlines the methods employed to investigate how word-medial intervocalic /t/ is realised across West Yorkshire. It describes how the /t/ tokens were selected and analysed, as well as how the statistical analysis was performed.

6.4.1. Token selection

Tokens of word-medial, intervocalic /t/ were selected from clearly articulated speech where there was no uncertainty as to what the intended target was. Any tokens produced in overlap or when the participant was laughing were disregarded, as were words that had been almost fully elided due to co-articulation. Tokens were only selected from content words, as opposed to function words, and contracted forms such as *gotta* were excluded, as were acronyms such as *CCTV*. For each participant, all suitable tokens were manually labelled in Praat (Boersma & Weenink, 2019) using a TextGrid. Labels included the word containing the intervocalic /t/ token and transcription of the token. A range of specific keywords related to the task occurred frequently in almost all of the participants' Task 1 recordings, such as *A-forty, Carter, hotel*, *Peter* and *scooter*. Appendix 2 contains a frequency table with all of the /t/ tokens that were produced by the 30 participants.

6.4.2. Auditory analysis

In line with many previous studies of /t/, an auditory analysis was undertaken in order to explore T-glottaling rates as well as how /t/ is realised more generally. Although variability in /t/ could have been explored by examining continuous acoustic aspects of the variable such as VOT, the categorical feature of glottalisation was of more interest in this study, as it is considered to be much more socially salient and likely to be influenced by regional variability and accommodation effects. All tokens were manually transcribed based on auditory perceptions, which were verified by visually inspecting the acoustic information available in the spectrogram and waveform. However, no acoustic measurements were taken. A subset of 20% of the sound files were also analysed by a second analyst, Dr Erica Gold, over multiple sessions. We initially discussed our findings and, in any instances where differences arose, we agreed on an approach for classifying less straightforward tokens that was consistently implemented going forward.

In cases where plosives were released with multiple bursts, the acoustic information relating to the initial burst was taken into account. No attempt was made to formally quantify and analyse the frequency of multiple released plosives, as it was not anticipated that this phenomenon would be regionally stratified or that speakers would accommodate in this respect. Previous studies that have examined plosives released with multiple bursts have indicated that a main determining factor in how many bursts occur is the place of articulation, with multiple bursts being more likely to occur in velar stops (Gráczi & Kohári, 2014). Furthermore, Gráczi & Kohári note that between-speaker differences may also be explained with reference to "the cavity sizes, tongue mass and intraoral and subglottal pressure characteristics" (2014, p. 158). For this reason, multiple bursts are generally considered an idiosyncratic feature as opposed to being specifically accent-related or dependent on the context of the interaction.

Within the WYRED dataset, the following variants were identified auditorily: [t], [t^{s}], [t͡s], [ʔ], [ʔ]* [ʔ͡t], [ɾ], [s], [k] and Ø. Figures 6.1-6.10 provide examples of each type of variant that were observed during the auditory analysis. The vast majority of tokens that were auditorily perceived as glottal variants did not exhibit the acoustic cues we might expect to find in the speech signal. For instance, there was no silent hold phase, or 'stop gap', which we would generally expect to see with all voiceless stops. Furthermore, there was rarely a clearly visible vertical striation to mark the plosive release. Instead, the percept of a glottal stop seemed to be prompted most often by a period of creaky voice. This was also found to be the case in Docherty & Foulkes' (1999) study of Newcastle speakers' glottal realisations, where it was noted that a number of scholars have claimed that glottal closure and creaky voice form a continuum (cf. Grice & Barry, 1991; Ladefoged & Maddieson, 1996). In Liverpool English, a typical pattern for [ʔ] was also a period of creaky voice (Clark & Watson, 2016, p. 37).

In the interest of transparency, tokens were coded as either [ʔ] or [ʔ]* depending on whether or not the glottal stop was visible in the acoustic signal, respectively. Figure 6.1 shows an example of /t/ being realised as [ʔ] where there is evidence of a glottal stop in the acoustic signal, whereas Figure 6.2 presents an example that was coded as [ʔ]*. Regardless of the acoustic differences in these two types of tokens, the auditory perception in both cases was that the underlying /t/ had been replaced with an auditorily distinct glottal stop [ʔ].

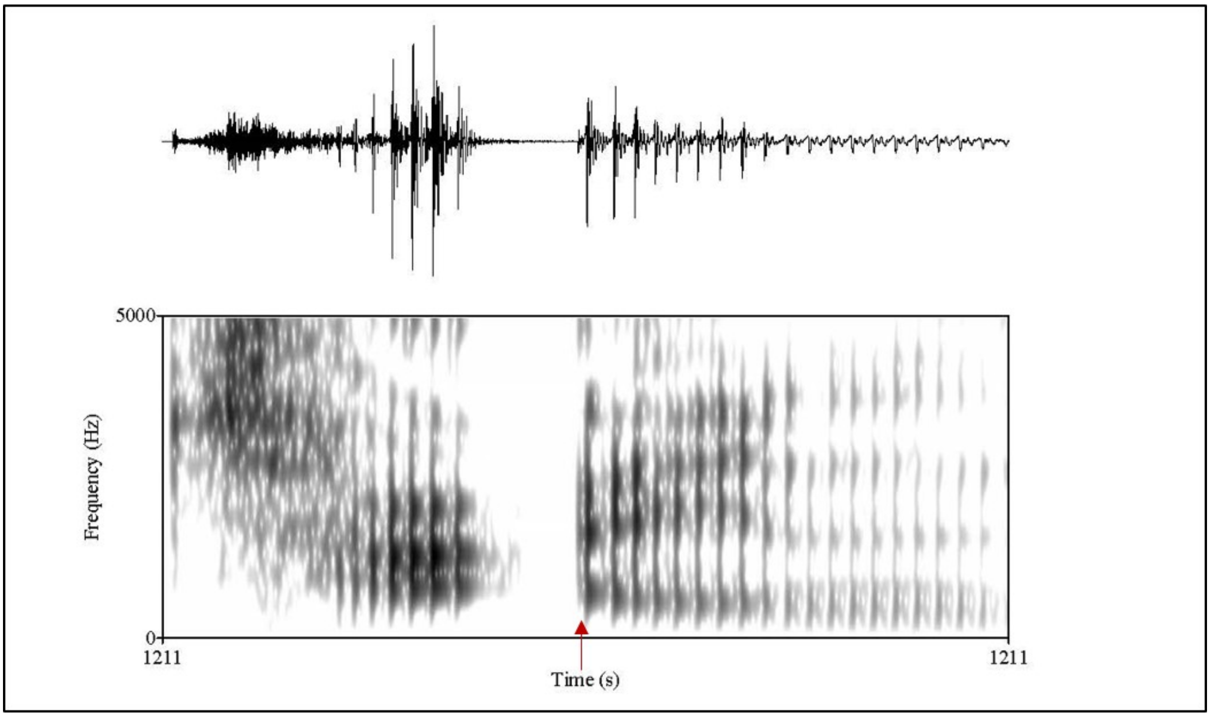


Figure 6.1. Spectrogram of *chatting* spoken by Wakefield participant #054, showing [ʔ] (Arrow indicates the glottal release).

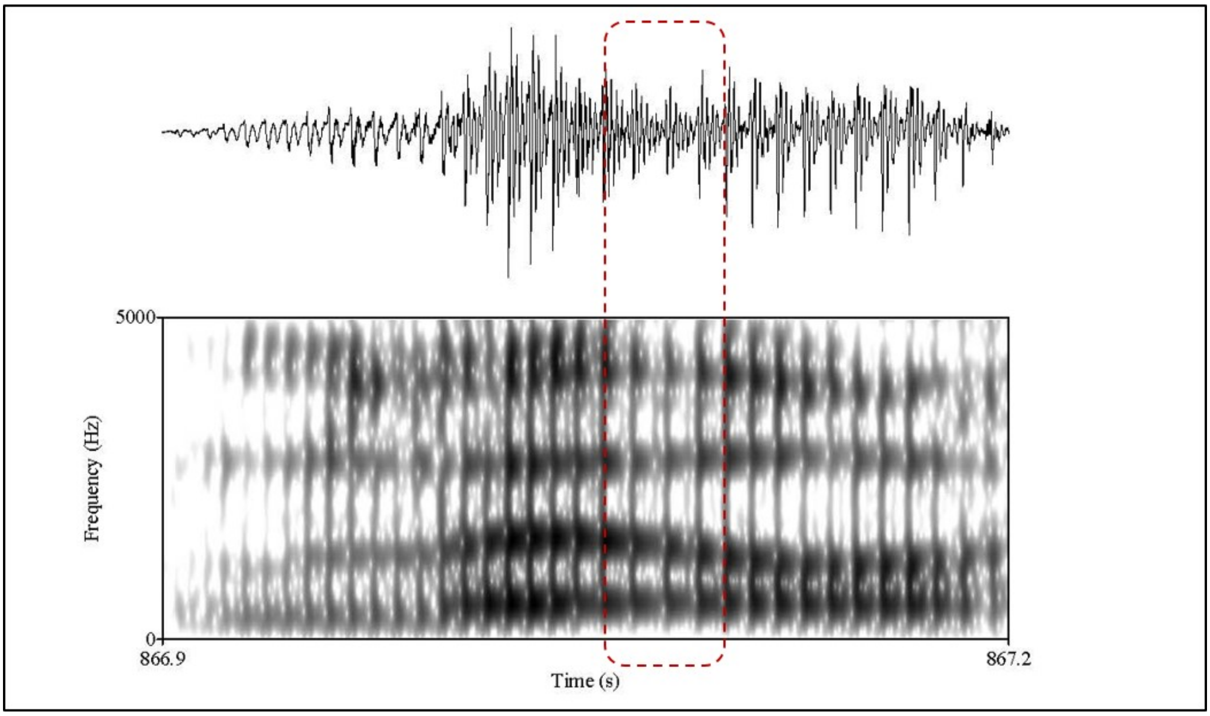


Figure 6.2. Spectrogram of *little* spoken by Bradford participant #031, perceived as a glottal variant despite the absence of 'stop gap' or an acoustic transient in the signal (Rectangle indicates portion of creaky voice where glottal is auditorily perceived).

Figure 6.3 provides an example of a glottalised variant [ʔ͡t] where there appears to be a double articulation involving both an oral closure and a glottal closure. Similar to many of the glottal tokens, the auditory perception of the glottal articulation is created by a period of creaky voice, which is then followed by a fully released [t]. These tokens are considered to be similar to the type of glottalised variants that are typically found in the Tyneside region (Docherty & Foulkes, 1999; Llamas, 2007; Watt & Milroy, 1999). Figure 6.4 provides an example of a 'canonical' [t] where the typical acoustic cues of a voiceless plosive can be seen. There is an onset phase at the end of the preceding vowel where the articulators come together to initiate the closure. This closure is then maintained throughout the hold phase for approximately 60 ms and then released, resulting in an acoustic transient. In contrast to this, Figure 6.5 shows an example of /t/ being fully elided in the production of the word *Peter*. In tokens such as this, the vowels preceding and following the underlying /t/ transition into one another without the presence of an oral or glottal closure between them.

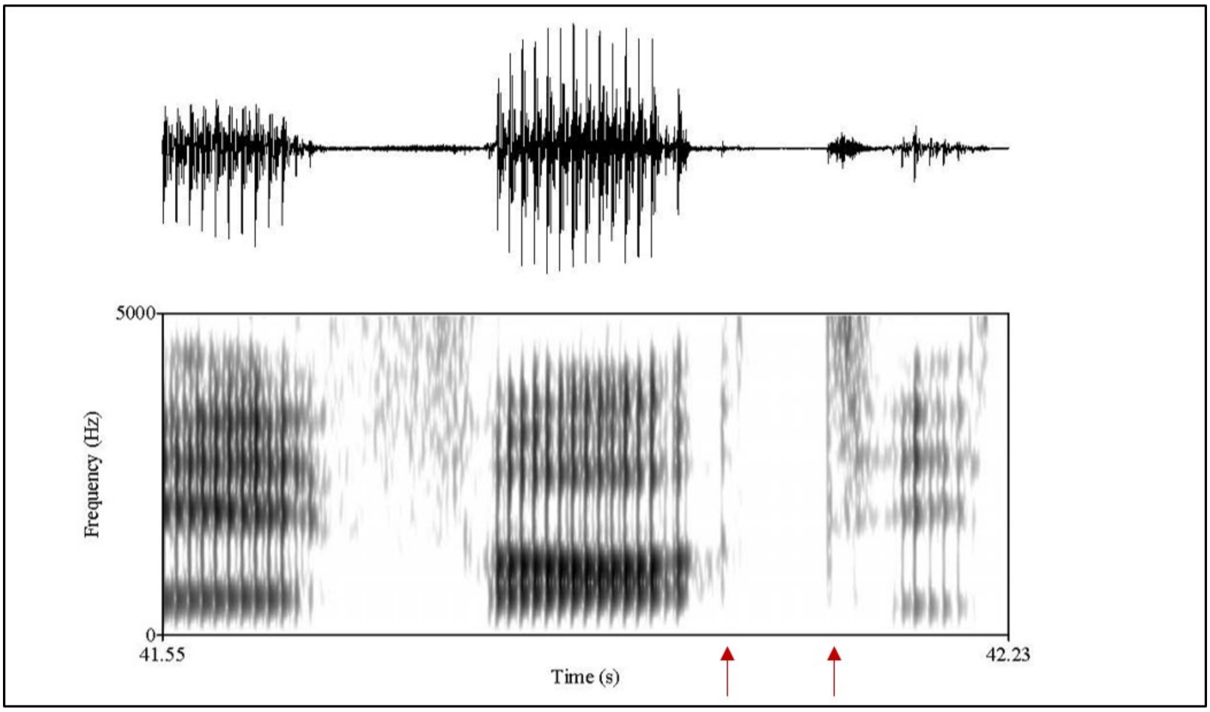


Figure 6.3. Spectrogram of *A-forty* spoken by Wakefield participant #054, showing [ʔ͡t] (First arrow marks the glottal release, the second arrow marks release of [t]).

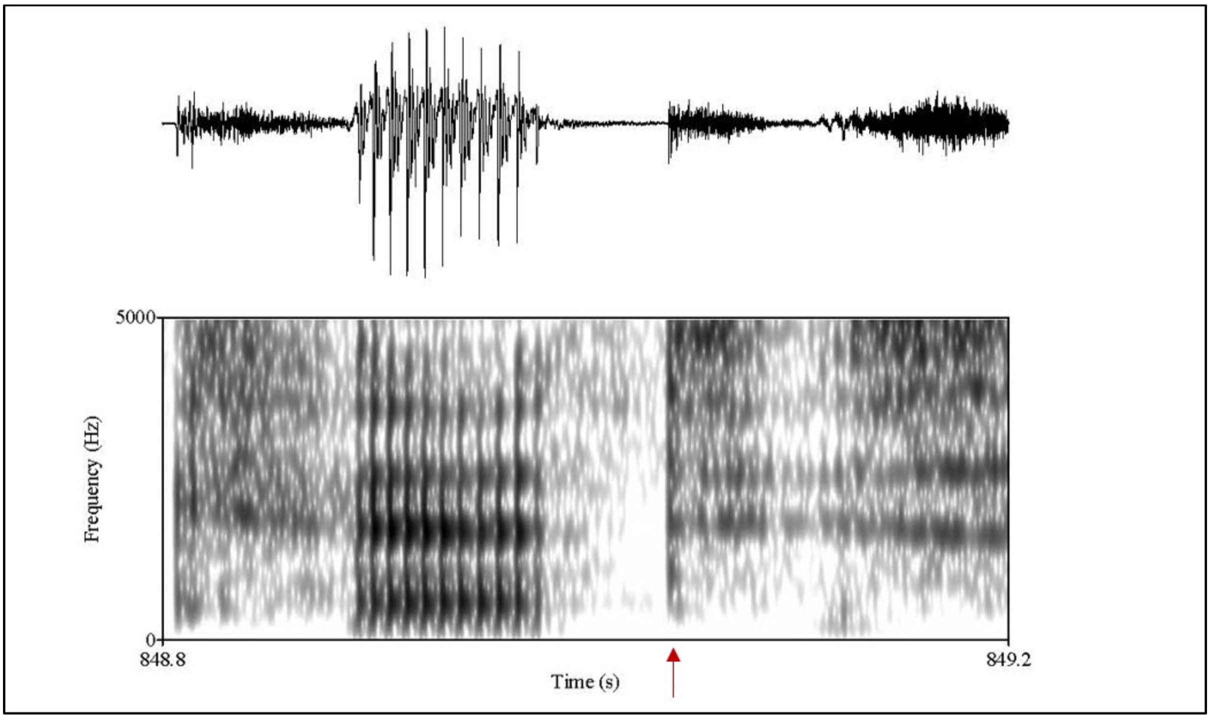


Figure 6.4. Spectrogram of *Curtis* spoken by Bradford participant #031, showing released 'canonical' [t] (Arrow marks release of [t]).

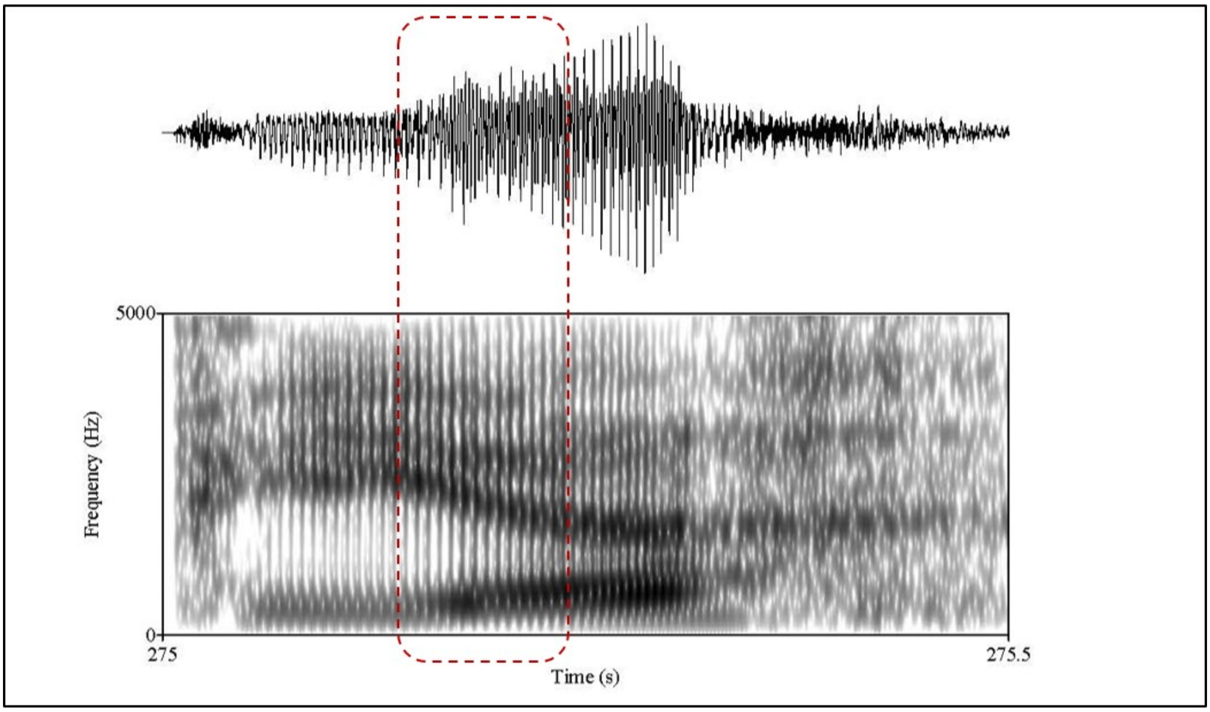


Figure 6.5. Spectrogram of *Peter* spoken by Wakefield participant #006, with fully elided /t/ (Rectangle marks smooth transition between the vowels [iː] and [ə]).

Figure 6.6 provides an example of a tapped realisation of /t/. In this token, there is a short closure period of approximately 20 ms, which is voiced all the way through. This is then followed by a transient, which corresponds to the release of the tongue after striking the alveolar ridge. Figure 6.7 presents the only example of /t/ being realised as [k] that was observed within the dataset. In line with the reported usage of this variant for /t/ in Manchester English (Baranowski & Turton, 2015, p. 305), this occurred before syllabic /l/ in the word *beetle*. In this specific token, the same acoustic cues that were observed in the canonical [t] tokens are present; however, in this case the oral closure is formed further back at the velum as opposed to the alveolar ridge.

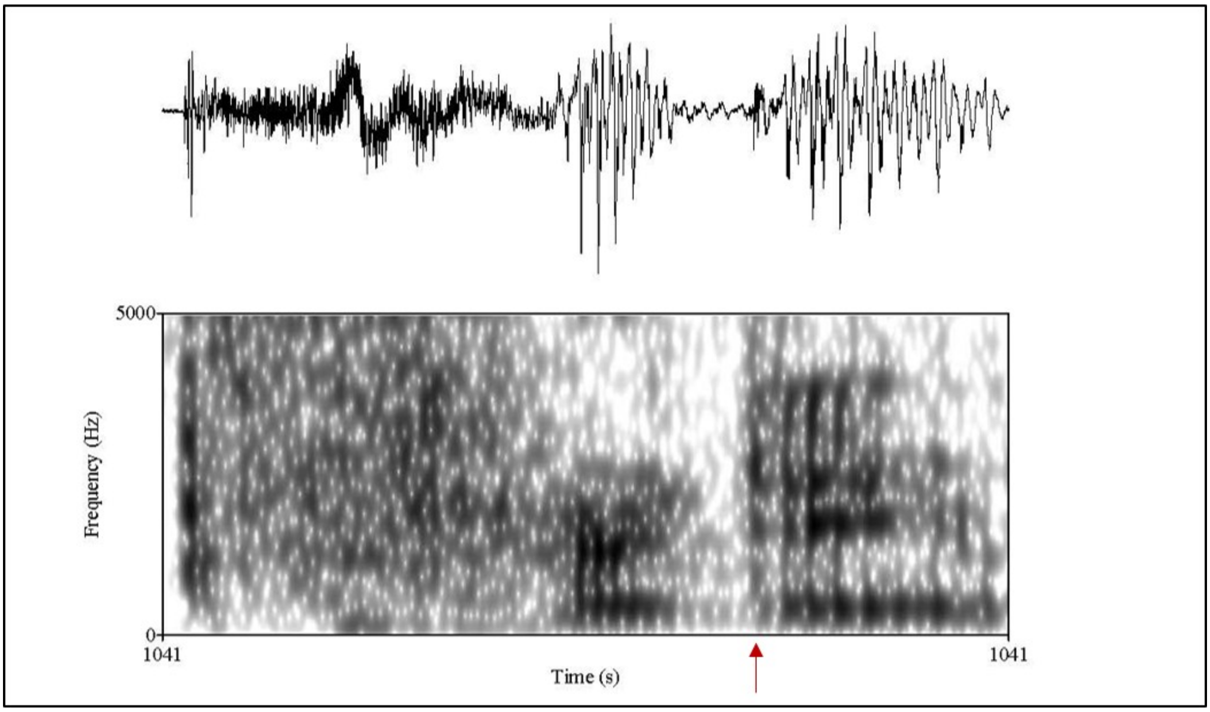


Figure 6.6 Spectrogram of *pretty* spoken by Kirklees participant #021, showing [ɾ] (Tap marked with arrow).

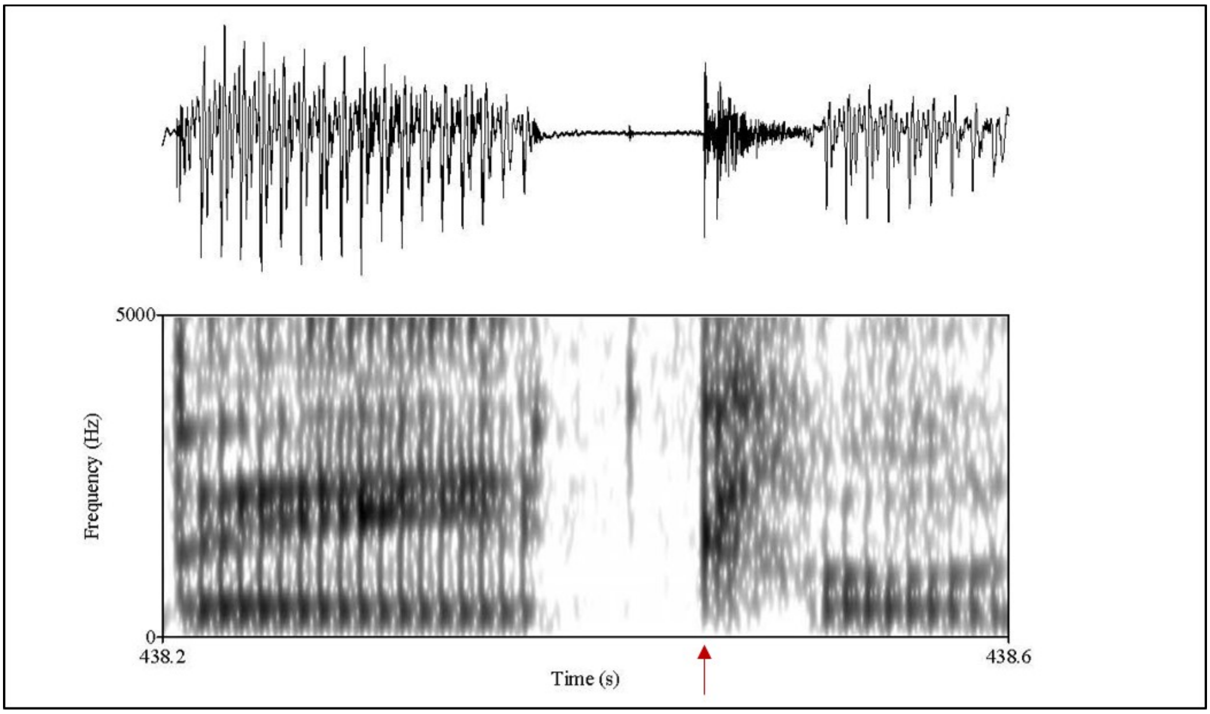


Figure 6.7. Spectrogram of *beetle* spoken by Kirklees participant #015, showing [k] (Arrow marks the release of [k]).

Figures 6.8-6.10 provide examples of /t/ being produced with varying degrees of fricative energy. The [tˢ] token displayed in Figure 6.8 is affricated as a result of being released with a marked degree of fricative noise. Tokens of this type are similar to those commonly found in Liverpool English (Hughes et al., 2012, p. 113). Figure 6.9 presents an example of /t/ produced with a double articulation involving a voiceless alveolar plosive and a voiceless fricative. In this example, there is a clear transient marking the release of the oral closure followed immediately by the onset of high frequency energy, which is maintained for approximately 125 ms. Figure 6.10 presents an example of /t/ being fully spirantised to an [s]-like quality, where no closure is evident in the acoustic signal. There were only three instances where /t/ was realised as [s] within the dataset.

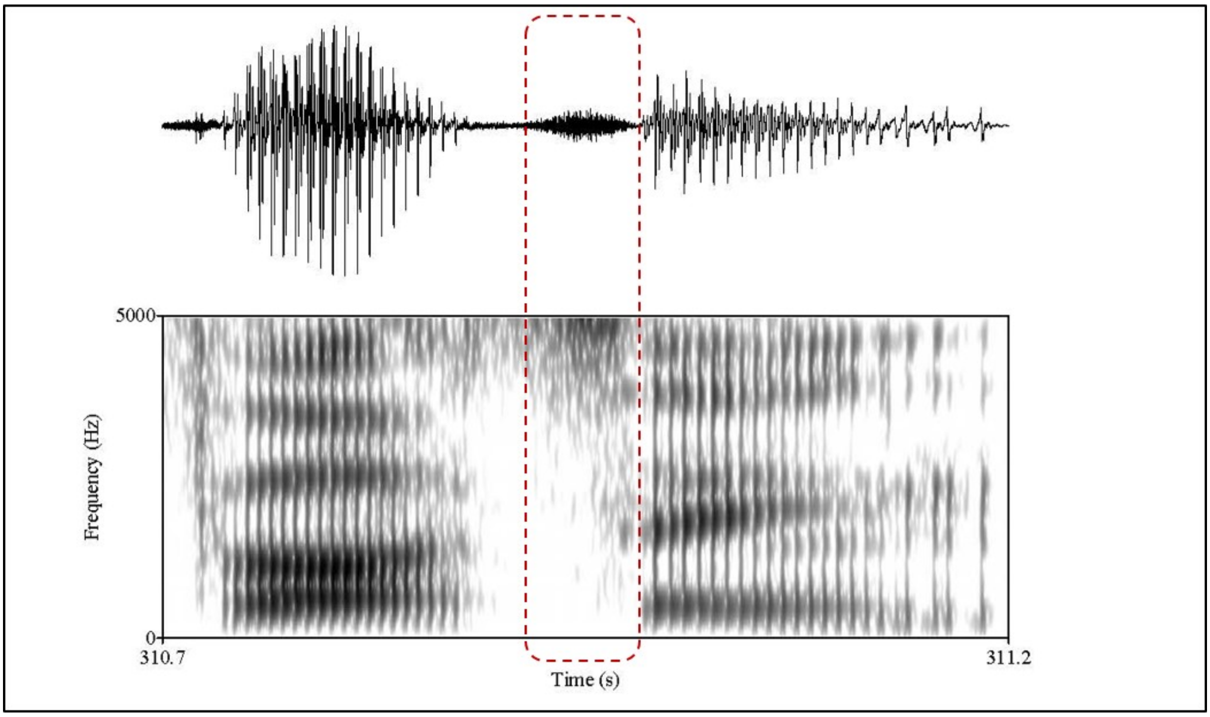


Figure 6.8. Spectrogram of *fourteen* spoken by Bradford participant #022, showing [tˢ] (Rectangle marks affrication).

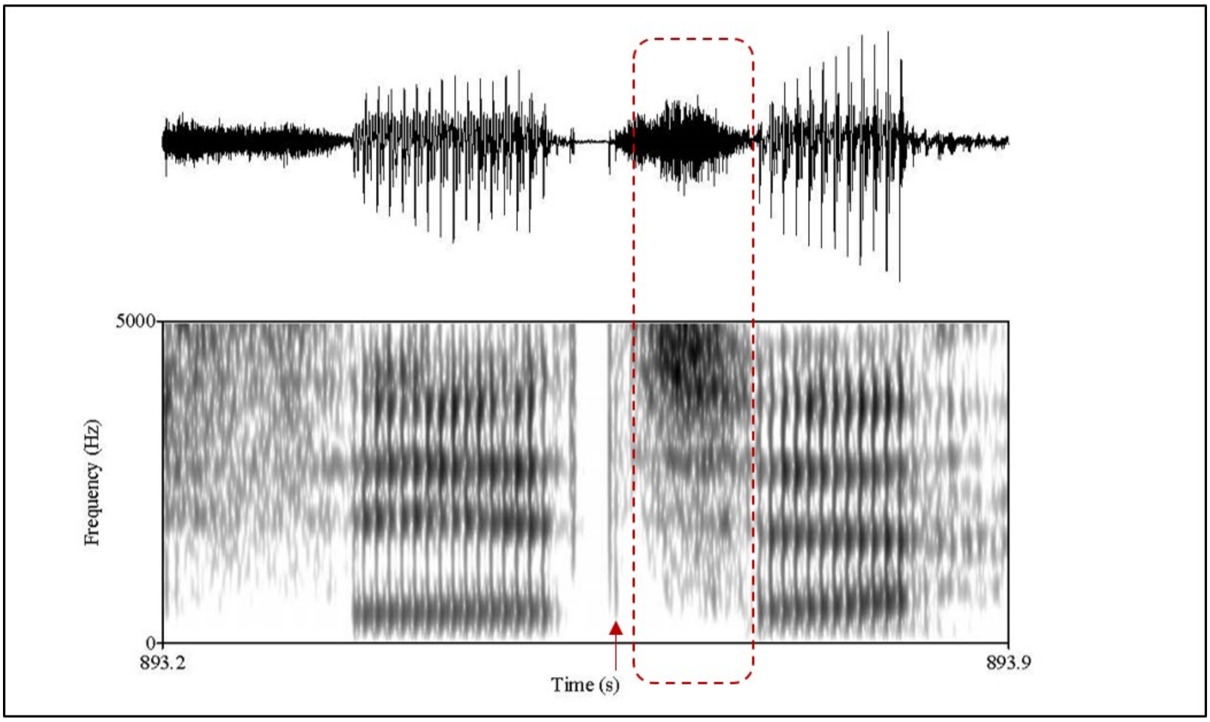


Figure 6.9 Spectrogram of *theatre* spoken by Bradford participant #040, showing [t͡s] (Arrow marks release of stop, rectangle marks fricative [s]).

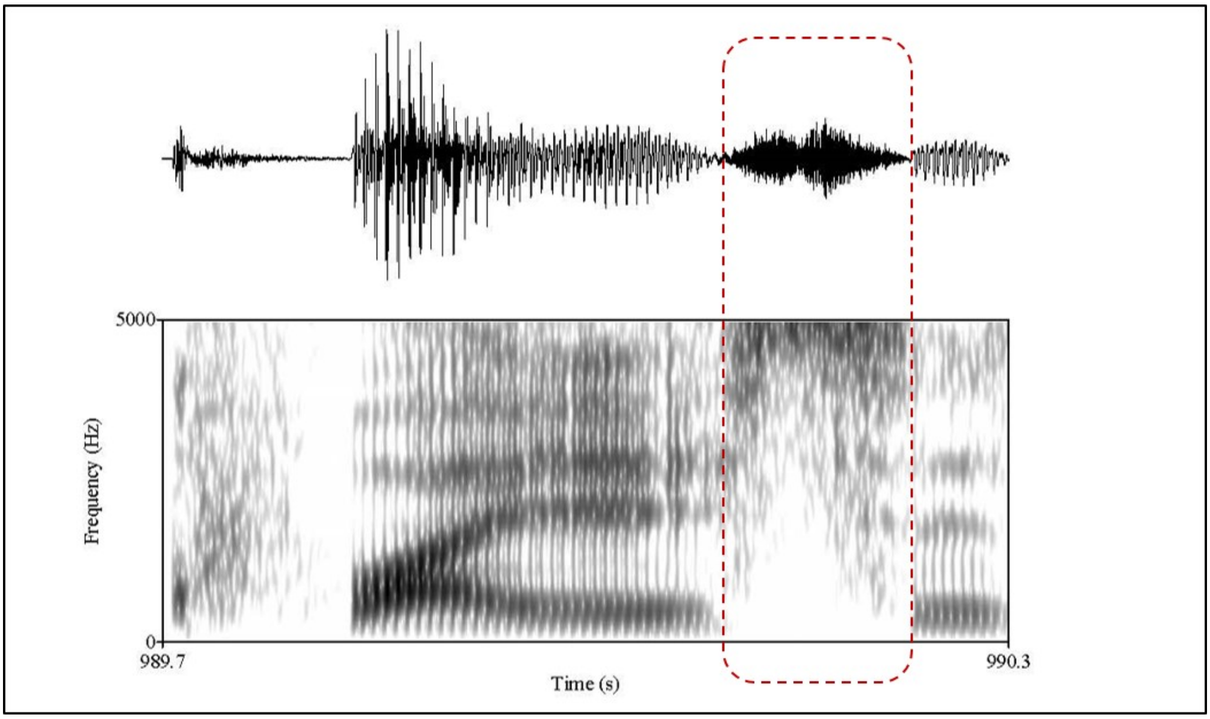


Figure 6.10. Spectrogram of *quieter* spoken by Wakefield participant #041, showing [s] (Fricative marked with rectangle).

6.4.3. Data processing

In total, 1599 intervocalic /t/ tokens were analysed across all participants and tasks. The total number of tokens analysed across tasks and boroughs are summarised in Table 6.4. Due to the relatively short length of the Task 4 recordings, significantly less tokens were available in Task 4 than in Tasks 1 and 3. The median number of tokens selected from Tasks 1, 3 and 4 were 29, 22 and 2 respectively. As there were so few tokens available in the Task 4 recordings, the Task 4 data was not included in any of the statistical analyses presented in this chapter.

Table 6.4. Number of /t/ tokens analysed, by borough and task.

	Number of /t/ tokens analysed			
Area	**Task 1**	**Task 3**	**Task 4**	**All Tasks**
Bradford	302	230	13	545
Kirklees	276	200	20	496
Wakefield	316	213	29	558
All 3 boroughs	894	643	62	1599

As previously mentioned in Section 6.1.3, syllabic position and following phonetic environment can influence the realisation of /t/, therefore care was taken to group tokens into appropriate categories, prior to conducting any analysis of the data. Table 6.5 provides a summary of the three phonetic environments that the word-medial, intervocalic /t/ tokens have been grouped into in this investigation. In the same way that Wells' (1982) lexical sets are used to represent particular vowel categories, the three separate keywords shown in Table 6.5 will be used from this point forward to refer to each category of tokens. Tokens were included in the HOTEL category if /t/ occurred in the onset of a stressed syllable, word-medially (i.e. where the following nucleus was more prominent than the preceding one). In any instances where the stress could theoretically be placed on more than one syllable (e.g. *eighteen* realised as either /ˈeɪtiːn/ or /eɪˈtiːn/) all potential variable tokens were checked auditorily before being coded accordingly. Tokens were assigned to the PRETTY category where /t/ appeared between two vowels in an unstressed position. Unstressed tokens occurring after a vowel and before a syllabic consonant were assigned to the LITTLE category.

Table 6.5. Phonetic environment of intervocalic /t/ tokens.

Keyword	Description
HOTEL	Foot-initial/stressed, intervocalic /t/
PRETTY	Foot-medial/unstressed, intervocalic /t/ (preceding a vowel)
LITTLE	Foot-medial/unstressed, intervocalic /t/ (preceding a syllabic consonant)

Table 6.6 provides a breakdown of the number of /t/ tokens analysed across each of the phonetic environments by borough and across West Yorkshire as a whole. Overall, the vast majority of tokens were from the PRETTY context (74.4%). 16.4% were from the HOTEL context and 9.2% were from the LITTLE context. The distribution of /t/ tokens across the three phonetic environments were broadly similar across the three boroughs.

Table 6.6. Number of /t/ tokens analysed, by borough and phonetic environment.

Area	Number of /t/ tokens analysed		
	HOTEL	**PRETTY**	**LITTLE**
Bradford	77	418	50
Kirklees	91	353	52
Wakefield	94	419	45
All 3 boroughs	262	1190	147

6.4.4. Statistical analysis

In addition to examining how word-medial, intervocalic /t/ is realised across West Yorkshire, a key motivation of this study was to establish if /t/ productions vary according to location (i.e. across the boroughs of Bradford, Kirklees and Wakefield) and if so, to what extent. In order to test this, a statistical analysis was carried out to examine the effects of borough, as well as the effects of phonetic environment and speaking task on T-glottaling rates. The justification for focusing on the amount the participants produced glottal variants of /t/, as opposed to affricated or tapped realisations of /t/ for instance, was based on an initial analysis that revealed that the variation in West Yorkshire is largely between [t] and glottal stop.

In many of the previous sociolinguistic studies that have investigated glottal forms of /t/, auditory analysis has been used to distinguish variants involving a glottal articulation, [ʔ] and/or [ʔ͡t], from non-glottal variants [t]. In line with this, a binary distinction between glottal versus non-glottal variants was drawn for the purposes of statistical analysis in the present study; with glottal tokens coded as 1 and all other tokens coded as 0. The glottal category included variants [ʔ], [ʔ]* and [ʔ͡t] whereas the non-glottal category included all of the other variants that occurred in the dataset. Therefore, for the purposes of the statistical analysis of T-glottaling rates, "T-glottaling" refers not only to instances where /t/ is replaced with a glottal stop, but also with glottalised variants. As previously noted, the Task 4 data was excluded from this analysis due to being so limited. Furthermore, HOTEL tokens were not included in this analysis, as tokens produced in this phonetic environment were not within the envelope of variation. This is because HOTEL tokens were never realised as a glottal variant in this dataset, even in the so-called –ee/-oo environments. This resulted in the 1282 remaining tokens being included in the statistical analyses.

R version 3.5.0 (R Core Team, 2018) and lme4 version 1.1-21 (Bates, Mächler, Bolker, & Walker, 2015) were used to perform a generalised mixed effects logistic regression analysis in order to test the effect of borough, task and phonetic environment on T-glottaling rates. A generalised mixed effects logistic regression analysis was deemed to be appropriate as this study involves a binary outcome variable (glottal versus non-glottal) as well as both fixed and random effects. As a fixed effect, BOROUGH was entered into the model, and this was treated as a categorical factor with three levels (Bradford, Kirklees and Wakefield, with Bradford as the reference level). The model also needed to account for any potential effects of contrasting speech styles across tasks and articulatory differences across the different phonetic environments from which /t/ tokens were collected. For this reason, TASK and PHONETIC ENVIRONMENT were also entered into the model as fixed effects, with both factors having two categorical levels: (Task 1 and Task 3, with the former as the reference level) and (PRETTY and LITTLE, with the former as the reference level), respectively.

As each participant produced multiple intervocalic /t/ tokens, these responses could not be regarded as being independent from one another. In order to deal with this inter-dependency, a random effect for PARTICIPANT was entered into the model. It also had to be taken into account that the effect of TASK may not be equal for all participants, due to the fact that accommodation behaviour was expected to be relatively speaker specific, and therefore by-participant random slopes added for the effect of TASK were included. The full model with /t/ productions as the dependent variable is presented below:

T-glottaling.model = glmer(/t/ ~ BOROUGH + TASK + CONTEXT + (1+TASK|PARTICIPANT), data=Data, family=binomial)

In this model, the fixed effects of BOROUGH, TASK and PHONETIC ENVIRONMENT (without interaction terms) were used to predict T-glottaling rates. As random effects, there were intercepts for participants, as well as by-participant random slopes for the effect of TASK. The model was fitted using maximum likelihood and in order to test the influence of each of the fixed effects on T-glottaling rates, p-values were obtained by likelihood ratio tests of the full model with all of the fixed effects against the model without each of the fixed effects in question.

A further model was also constructed which was the same as the one above, except that it included interaction terms between each of the fixed effects. The purpose of this model was to be able to evaluate whether any interactions that may exist between the fixed effects had a significant effect on T-glottaling rates. This second model with /t/ productions as the dependent variable is presented below:

T-glottaling.model.int = glmer(/t/ ~ BOROUGH*TASK + BOROUGH*CONTEXT + TASK*CONTEXT + (1+TASK|PARTICIPANT), data=Data, family=binomial)

6.5. Results

This section presents quantitative results based on the auditory analysis of word-medial, intervocalic /t/. Firstly, an overview is presented of the distribution of variants observed across West Yorkshire. Secondly, the results of the generalised linear mixed effects analysis are provided in order to examine the extent to which T-glottaling rates differ across West Yorkshire boroughs, as well as how they are influenced by the phonetic environment and the speaking task.

6.5.1. /t/ in West Yorkshire

In total, 1599 tokens of intervocalic /t/ were analysed within the speech of 30 participants completing three separate speaking tasks and a range of variants were observed in the data. The distribution of these variants is presented in Table 6.7, with variants ordered from most to least frequently use. Overall, 93.4% of tokens are realised as either a glottal variant (represented as [ʔ] or [ʔ]*) or a fully released voiceless alveolar plosive. In a small proportion of tokens, /t/ is fully elided without any glottal closure or period of creaky voice. There are also some affricated forms, however, these tokens account for less than 2% of the data. There are 10 examples or fewer of all other /t/ variants.

Table 6.7. Overall distribution of all /t/ variants.

Variant	[ʔ]*	[t]	[ʔ]	Ø	[tˢ]	[ɾ]	[t͡s]	[ʔ͡t]	[s]	[k]
%	45.2	37.2	11.0	3.1	1.7	0.6	0.5	0.4	0.2	0.1
N	723	595	176	50	27	10	8	6	3	1

Figure 6.11 presents the proportional distribution of the /t/ variants across the boroughs of Bradford, Kirklees and Wakefield. It can be seen that the main patterns are largely consistent across boroughs with the majority of tokens being either glottal variants or [t]. It would appear that Kirklees has a slightly higher proportion of glottal variants than Bradford and Wakefield. Bradford appears to have the highest proportion of affricated [tˢ] tokens (4% of all tokens produced by Bradford speakers, compared to 0.6% in Kirklees and 0.4% in Wakefield). However, these findings do not take into account any differences in the distribution of tokens across phonetic environments or speaking task across each of the three boroughs.

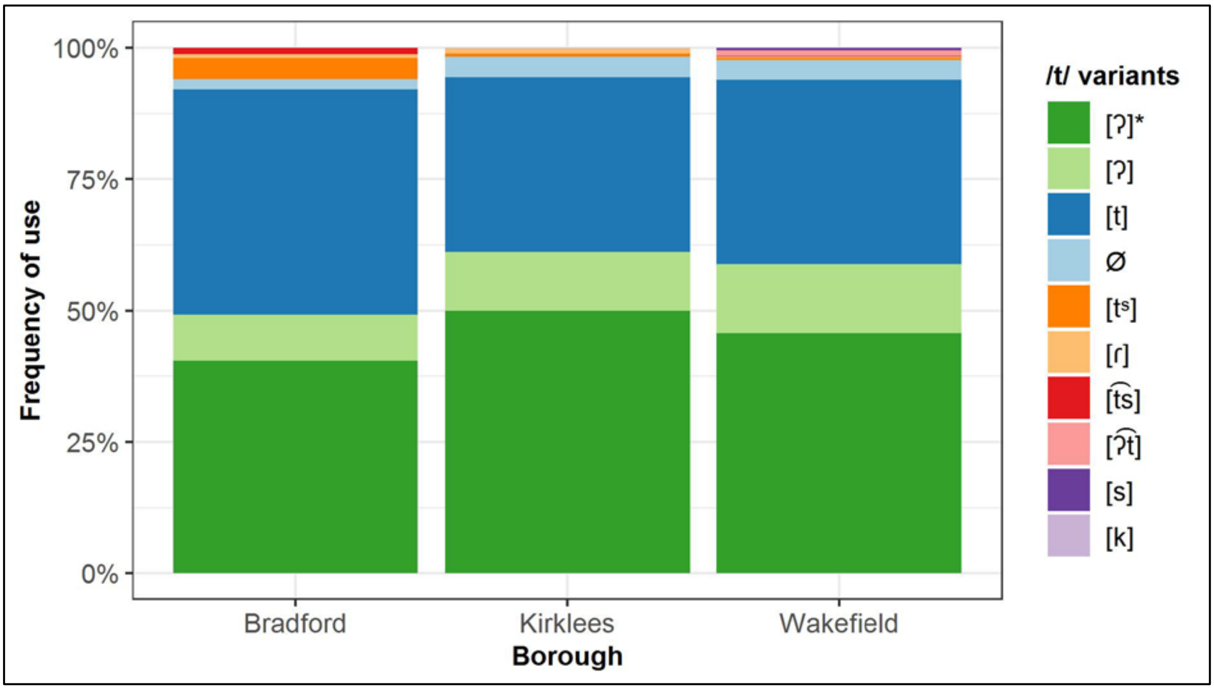


Figure 6.11. Distribution of /t/ variants across boroughs.

Figure 6.12 presents the proportional distribution of the various /t/ variants across the three phonetic environments considered in this investigation. It can be seen that in the HOTEL context, participants across all three boroughs only use the variants [t] and [tˢ]. As there are no instances of /t/ being realised as [ʔ] in a foot-initial/stressed position, this could indicate that T-glottaling is blocked for West Yorkshire speakers in this context, as it is in South East London English (Tollfree, 1999). In the PRETTY context, [ʔ]* is the most common variant followed by [t] then [ʔ]. Overall, glottal variants account for 65% of the PRETTY tokens. There are only 10 instances where /t/ is realised as the rhotic segment [ɾ], all of which occur in the PRETTY context. Regarding the LITTLE context, [ʔ]* is the most common variant followed by [ʔ] then [t]. Overall, glottal variants account for 90% of the LITTLE tokens. In addition to these variants, there is one example each of [s], [k] and full elision of /t/ to Ø in this context.

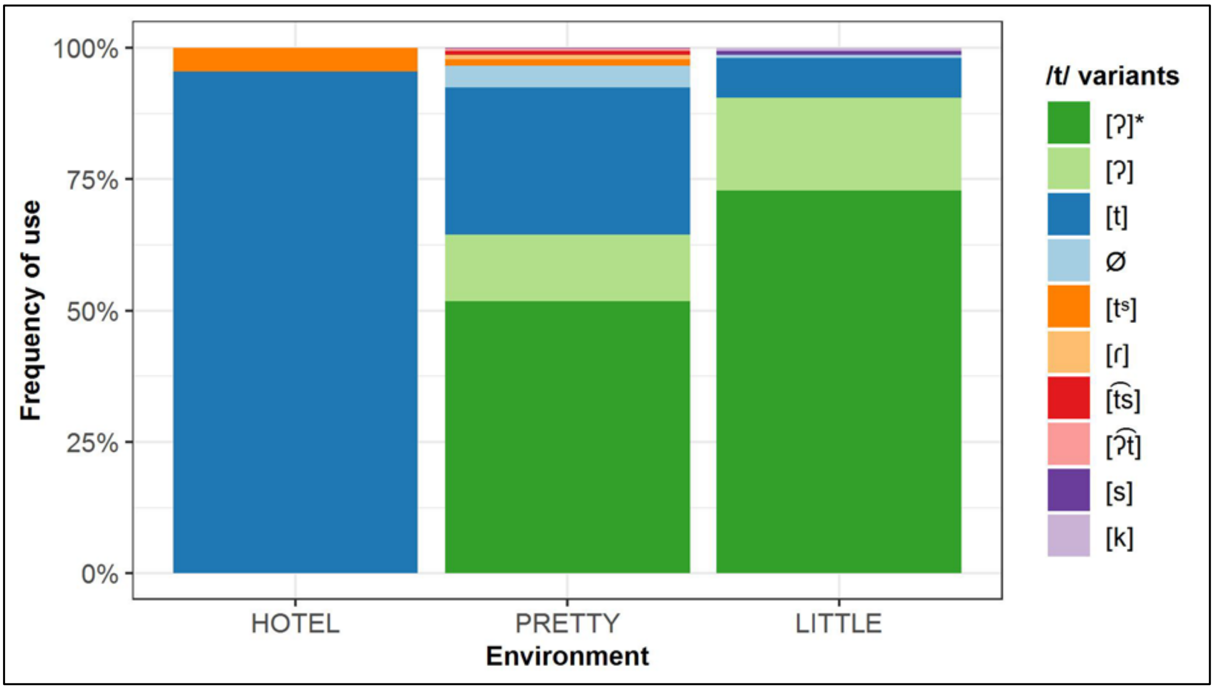


Figure 6.12. Distribution of /t/ variants across phonetic environments.

Figure 6.13 illustrates how the different /t/ variants are distributed across the boroughs of Bradford, Kirklees and Wakefield for each of the separate phonetic environments. In the HOTEL context, we can see that for all areas [t] is the most common variant with the only other variant being [tˢ]. Proportionately, the affricated variant is used most often in this context by Bradford speakers (10% of the time), followed by Kirklees (3% of the time) then Wakefield (1% of the time). In the PRETTY context, the distribution of the three most frequent variants are relatively similar across the three boroughs, although it can be seen that the Bradford participants have a higher incidence of [t] than Kirklees and Wakefield. All but one of the affricated tokens that occur in the PRETTY context are produced by participants from Bradford and all of the tokens produced with a glottalised variant [ʔ͡t] are produced by Wakefield participants. The tapped realisations of /t/ in the PRETTY context are distributed across all three boroughs. Finally, in the LITTLE context, the vast majority of tokens are glottal stops across the three boroughs (84% in Bradford, 94% in Kirklees, 93% in Wakefield). Similar to the PRETTY context, Bradford participants have a higher proportion of [t] tokens in the LITTLE context than Kirklees or Wakefield participants.

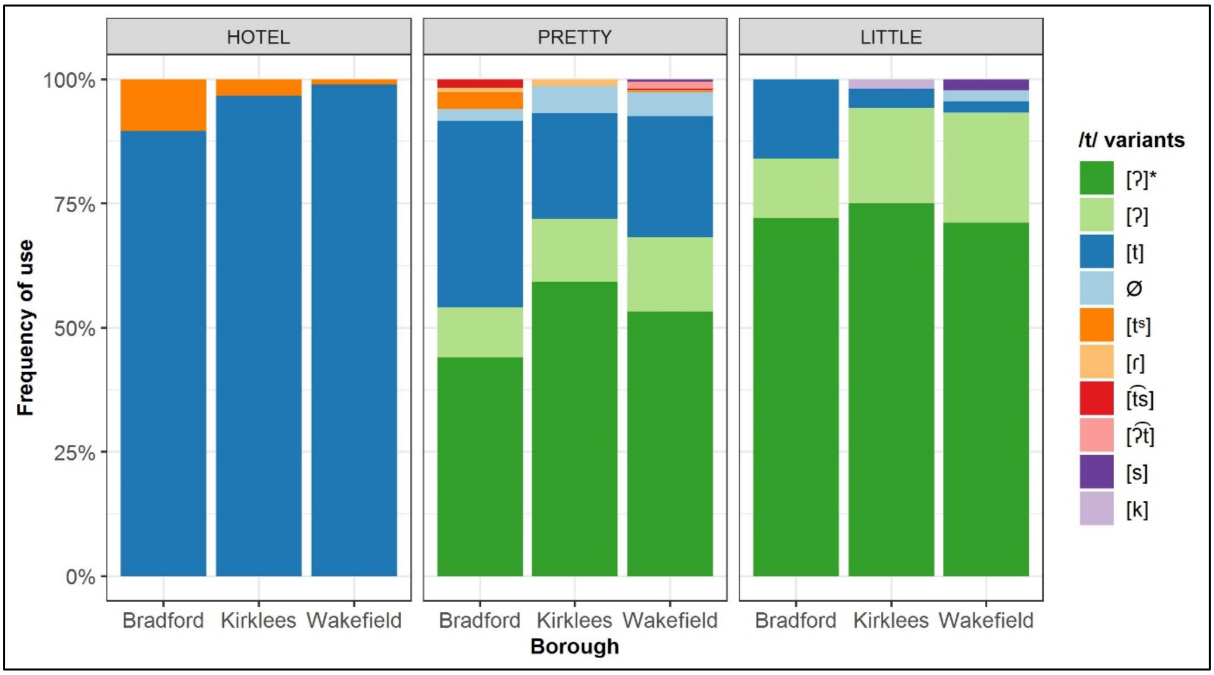


Figure 6.13. Distribution of /t/ variants by borough and phonetic environment.

Although it is not possible to discuss the findings of each of the 30 participants in detail, Appendix 5 provides bar charts to visualise the distribution of each participant's /t/ variants across each of the three phonetic environments and tasks. In these graphs it can be seen that many of the minority variants are only produced by a small number of specific individuals.

6.5.2. Generalised mixed effects logistic regression results

In order to test if /t/ productions were significantly different across boroughs, a statistical analysis was carried out to examine the effects of borough, as well as the effects of phonetic environment and speaking task on T-glottaling rates. The findings of the statistical analysis, revealed that the phonetic environment was the strongest predictor of T-glottaling in the model, with task also having a strong effect. Table 6.8 shows the results of the model of best fit, as determined by a generalised mixed effects logistic regression analysis, based on 1282 observations of word medial, intervocalic /t/. This model did not include any interaction terms. Any positive estimates in the regression coefficients indicated more use of T-glottaling within that category, and negative estimates meant the category was less likely to be realised as a glottal variant when compared with the reference level intercept. P-values were obtained separately by way of model comparison using likelihood ratio tests. Figure 6.14 presents the average glottal rates across the 30 participants, broken down according to borough, task and phonetic environment.

Table 6.8. Coefficients of a mixed-effects logistic regression model of T-glottaling, with random intercepts for PARTICIPANT (SD = 1.432) and by-participant random slopes for TASK (SD = 1.182).

	Estimate	Std. Error	z-value	Pr (>|z|)
Intercept (Bradford, Task 1, PRETTY)	-0.177	0.385	-0.458	0.647
Kirklees	0.764	0.468	1.634	0.102
Wakefield	0.487	0.451	1.080	0.280
Task 3	1.210	0.272	4.453	<0.0001
LITTLE	2.167	0.332	6.533	<0.0001

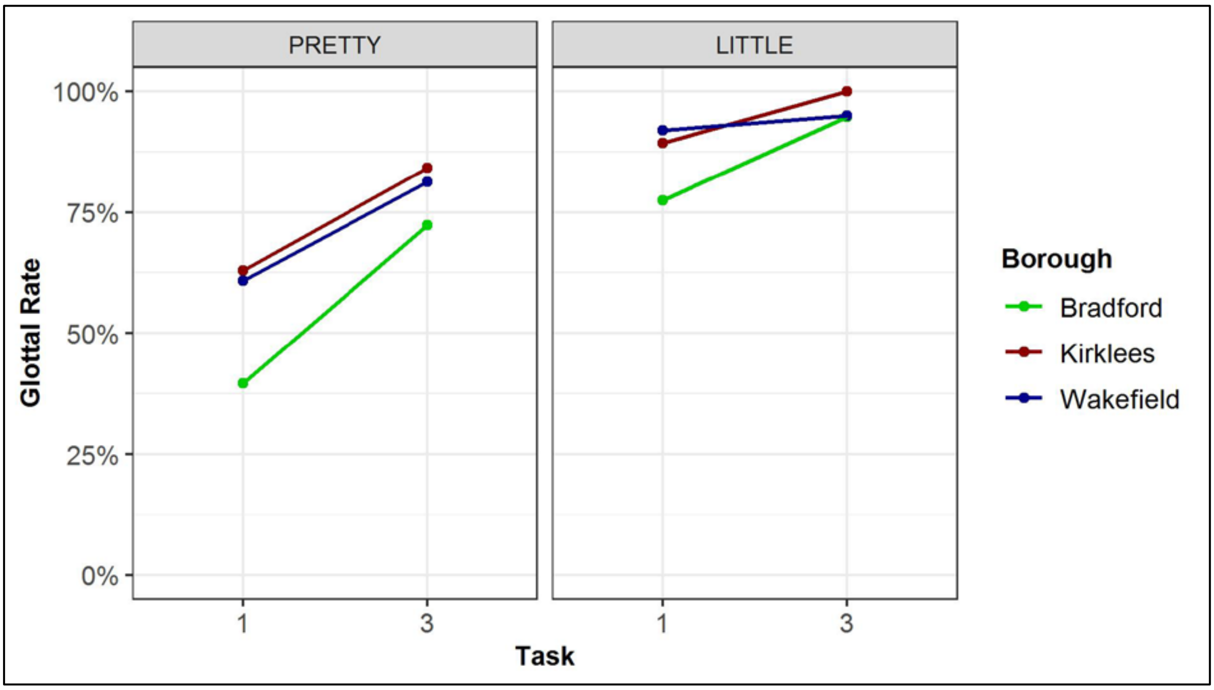


Figure 6.14. Glottal rates across borough, task and phonetic environment.

6.5.2.1 Influence of BOROUGH

In order to test the influence of the borough on T-glottaling rates, p-values were obtained by likelihood ratio tests of the full model, described in Section 6.4.4, against the model without the fixed effect of BOROUGH. This analysis showed that the borough did not have a statistically significant effect on T-glottaling (χ^2 (2) = 2.6398, p=0.267). This finding is also reflected in Figure 6.14 where it can be seen that particularly the boroughs of Kirklees and Wakefield appear to behave very similarly in respect of glottal rates. The Bradford participants generally tend to have a slightly lower T-glottaling rate than the other boroughs, however, this difference is relatively small, especially in the LITTLE context.

6.5.2.2 Influence of PHONETIC ENVIRONMENT

To explore the influence of the phonetic environment on T-glottaling, p-values were obtained by likelihood ratio tests of the full model against the model without the fixed effect of PHONETIC ENVIRONMENT. This analysis revealed that the phonetic environment had a statistically significant effect on T-glottaling (χ2 (1) = 58.795, p <0.0001), with glottal variants being most frequent in the LITTLE context (as can be seen from the positive estimate in Table 6.8). This finding is in line with previous descriptions of the linguistic patterning of T-glottaling in other accents. Figure 6.14 reflects this trend and also shows that, based on the average values in both the PRETTY and LITTLE contexts, glottal variants occurred more often than the non-glottal variants. The differences across the PRETTY and LITTLE contexts are also exemplified in the boxplot in Figure 6.15. In this plot, the range of glottal rates across each of the participants within the three West Yorkshire boroughs can be seen, broken down by phonetic environment. Please note that the darker the data points are, the more participants' glottal rates are represented. This graph demonstrates that the glottal rates were more spread out across participants from Bradford than they were for the Kirklees and Wakefield participants.

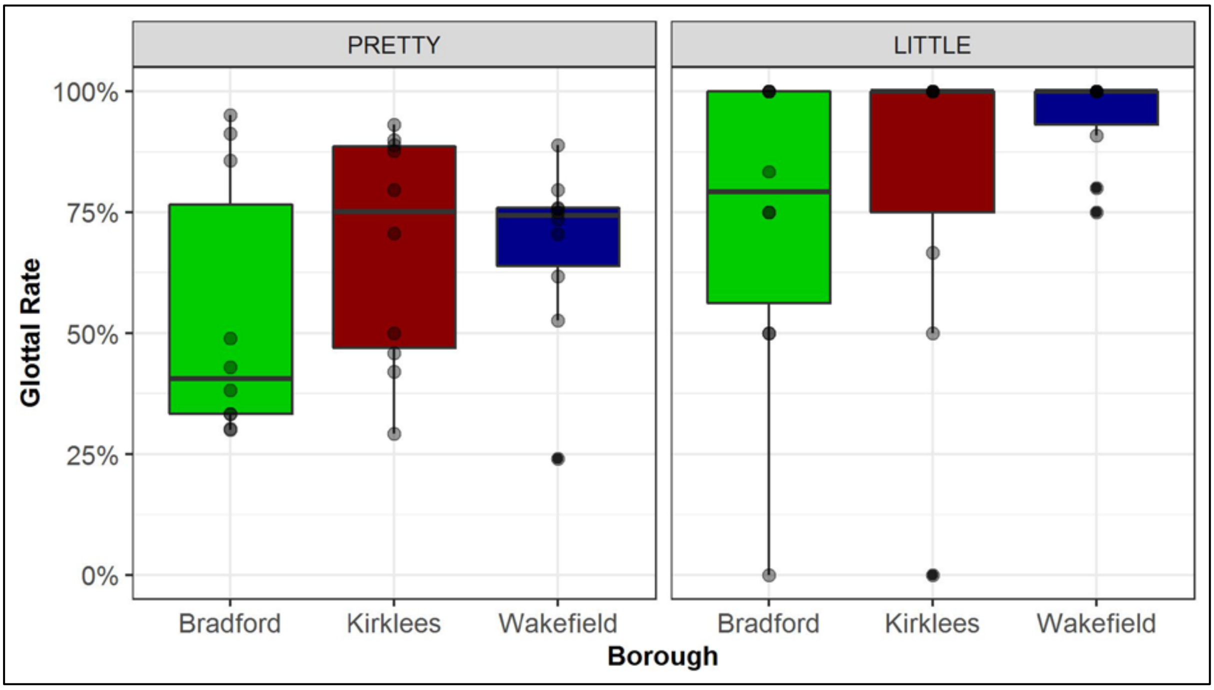


Figure 6.15. Glottal rates of all participants across boroughs, by phonetic environment.

6.5.2.3 Influence of TASK

To test the influence of the speaking task on T-glottaling, p-values were obtained by likelihood ratio tests of the full model, against the model without the fixed effect of TASK. This analysis revealed that the task had a statistically significant effect on T-glottaling (χ^2 (1) = 15.128, p <0.0001), with the glottal variant being more frequent in the casual paired conversation (Task 3) than in the mock police interview (Task 1). Again, this trend can be seen in Figure 6.14. The differences across the two tasks are also exemplified in the boxplot in Figure 6.16 where the range of glottal rates across each of the participants within the three West Yorkshire boroughs are broken down by task. This graph demonstrates that the glottal rates were generally higher in Task 3 than in Task 1, and again that there was some variability across borough - although this was not significant. The effects of speaking task are explored in much more detail in the following chapter as part of an investigation into accommodation effects in word-medial, intervocalic /t/. For this reason, further interpretation of this specific finding is reserved for the time being.

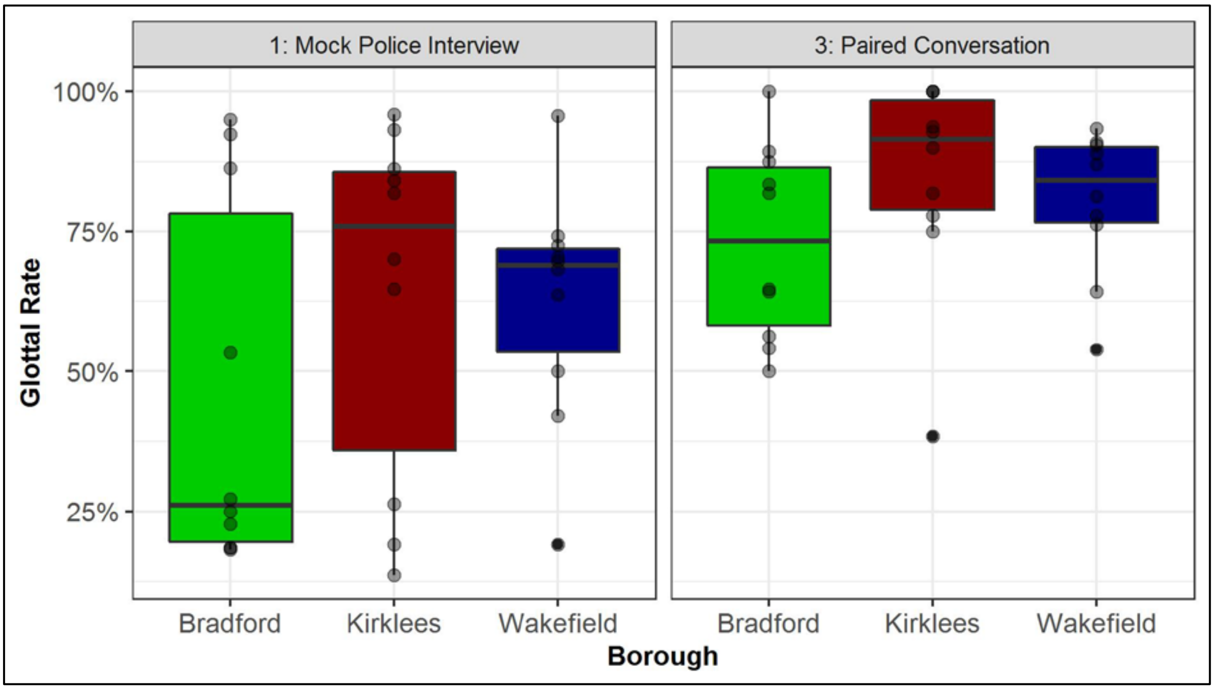


Figure 6.16. Glottal rates of all participants across boroughs, by task.

A further set of model comparisons were conducted using the second model that was described in Section 6.4.4 which included interactions between each of the fixed effects. Overall, it was determined that the model fit was reduced when interactions were included and the model failed to converge. This was potentially due to the model being overfit, most likely because of the imbalanced nature of the data across the separate cells. Despite this, the main patterns described above still held true whereby there were significant effects of both phonetic environment and speaking task on T-glottaling but no significant influence of borough. P-values were also obtained by likelihood ratio test of the full model against the model without each of the two-way interactions (BOROUGH*TASK/ BOROUGH*CONTEXT/ TASK*CONTEXT). In all cases, there were no significant effects of the interactions on T-glottaling rates and therefore none of the interactions appeared to contribute to the overall model fit. These results are reflected in the coefficients of the mixed-effects logistic regression model presented in Table 6.9.

Table 6.9. Coefficients of a mixed-effects logistic regression model of T-glottaling , with random intercepts for PARTICIPANT (SD = 1.422) and by-participant random slopes for TASK (SD = 1.164).

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept (Bradford, Task 1, PRETTY)	-0.374	0.483	-0.775	0.439
Kirklees	1.012	0.684	1.479	0.139
Wakefield	0.834	0.677	1.231	0.218
Task 3	1.442	0.451	3.202	0.001
LITTLE	2.255	0.490	4.597	<0.0001
Kirklees: Task 3	-0.291	0.660	-0.441	0.659
Wakefield: Task 3	-0.432	0.635	-0.680	0.497
Kirklees: LITTLE	-0.499	0.832	-0.600	0.549
Wakefield: LITTLE	-0.077	0.796	-0.096	0.923
Task 3: LITTLE	0.204	0.831	0.246	0.806

6.6. Discussion

In this section, the results of both of the research questions posed in Section 6.2 are discussed in detail and considered in relation to previous findings in the literature. Subsequently, the implications of the findings of this study for researchers in the fields of forensic speech science and sociophonetics more generally, are described.

6.6.1. Summary of findings

The results of this investigation revealed that throughout all three boroughs, T-glottaling is common in certain, but not all, word-medial intervocalic contexts. Participants typically tend to vary between [ʔ] and [t] in unstressed positions, with a preference for the glottal variant. T-glottaling occurs significantly more often in the LITTLE context than in the PRETTY context, whereas glottal stops are never found in the HOTEL context in this dataset. The fact that T-glottaling appears to be blocked when /t/ occurs in a stressed position suggests that T-glottaling is not in as advanced a stage in this speech community as it is in places such as

London (Harris & Kaye, 1990), Buckie (Smith & Holmes-Elliott, 2017) and Manchester (Baranowski & Turton, 2015).

Although T-glottaling does not appear to occur in the HOTEL context in West Yorkshire, there was evidence of a milder form of lenition taking place in the form of the affricated variant [tˢ], similar to those commonly reported in Liverpool English. As the HOTEL context would generally be considered a strong position where lenition is less likely to occur, this finding could potentially be taken as evidence of the next stage of the sound change. It is possible that this variant could have spread from another region such as Liverpool, where lenited forms are very typical, or it could be the case that they are a result of more idiosyncratic variation. It is interesting to note that although the [tˢ] variant is produced by at least one participant from each of the three boroughs, 81% of these tokens were produced by seven participants from Bradford. However, it is acknowledged that the affricated tokens only accounted for 4.6% of all HOTEL tokens and therefore this variant is very much in the minority. In the absence of more data, it is difficult to determine whether the use of the affricated forms are more regionally stratified than the glottal variants.

Taking into account the previous findings of Broadbent (2008), Hughes et al. (2012) and Petyt (1985), the results of this study indicate that T-glottaling has increased in West Yorkshire and may still be on the rise. This is perhaps unsurprising when considering that T-glottaling has a reputation for being one of the most rapidly spreading phonetic features across the UK (Smith & Holmes-Elliott, 2017). It would appear that West Yorkshire is participating in the consonantal change, in line with other nearby Northern cities such as Manchester (Baranowski & Turton, 2015), Sheffield (Stoddart et al., 1999) and Hull (Williams & Kerswill, 1999). In addition to the prevalence of T-glottaling in West Yorkshire, the relative absence of *t*-to-*r* for intervocalic /t/ in this area is also worth discussing further.

In total, there were only 10 instances where /t/ was realised as a rhotic variant. On these few occasions where the *t*-to-*r* rule appears to have applied, the variant was realised as [ɾ] rather than [ɹ], which may be expected as /ɹ/ is often realised as a tap in West Yorkshire (Hughes et al., 2012, p. 105). These variants mainly occurred after a short vowel, however, there were some tokens present in the dataset in which [ɾ] occurred after a long vowel, such as *forty* [fɔːɾɪ]. One explanation for this may be related to that fact that [ɾ] has more flexible lexical constraints on its distribution than [ɹ] (Clark & Watson, 2011). Overall, the findings of this investigation would support the theory proposed by Broadbent (2008) that as T-glottaling gradually becomes more popular, application of the *t*-to-*r* rule will simultaneously become less so. Tapped variants of /t/ accounted for less than 1% of the data overall.

Observations of the overall distributions of /t/ variants across the boroughs of Bradford, Kirklees and Wakefield indicate that participants from the three boroughs broadly pronounce word-medial, intervocalic /t/ in similar ways. As stated above, West Yorkshire participants typically tend to vary between [ʔ] and [t] in unstressed positions, with a preference for the glottal variant. When /t/ is in a foot-initial, stressed position it is realised as [t] except in a small number of cases where it is affricated. Although some individuals deviate from these variants occasionally, it would appear that [ʔ] and [t] are the most commonly used variants.

The statistical analysis of T-glottaling rates corroborated these observations and revealed that there were no significant differences in glottal rates across the three boroughs. This finding is in contrast to that of Petyt's (1985) investigation of /t/ in the areas of Bradford, Halifax and Huddersfield (Kirklees), in which Bradford was said to be leading the change towards an increased use of T-glottaling with this feature diffusing out to the neighbouring towns. The fact that Bradford no longer appears to have higher rates of [ʔ] than Kirklees (or Wakefield) indicates that the distribution of [ʔ] has changed over time in the West Yorkshire region and that the change that was in progress in 1985 is now complete, or at least close to completion.

Results of the statistical analysis also corroborated the preliminary observation that there were significant differences in T-glottaling rates across the PRETTY and LITTLE phonetic environments. The glottal variant was used much more frequently when /t/ preceded a syllabic consonant than when it preceded a vowel. This finding was in line with previous literature relating to the linguistic constraints on T-glottaling (Hughes et al., 2012, p. 67; Mathisen, 1999, p. 116; Watt & Milroy, 1999, p. 30). The statistical analysis also revealed that the influence of the speaking task also significantly contributed to the model fit, with T-glottaling being more likely to occur during the casual paired conversation than during the mock police interview task. This finding would suggest that participants tended to style-shift with respect to word-medial, intervocalic /t/ resulting in the non-standard variant being more likely to occur in the least formal context. This pattern is also considered to be in line with previous findings relating to the effects of speaking style on /t/ productions (Baranowski & Turton, 2015; Stuart-Smith, 1999b). This factor is addressed in more detail in the following chapter as part of an analysis of how /t/ is influenced by speech accommodation effects.

The findings of this study suggest that /t/ is not used by these West Yorkshire speakers to index any kind of local level identity specific to an individual borough. This may be because T-glottaling is generally considered a socially salient feature that is present throughout many different parts of the UK. It is clear that people from this area are aware of this feature in their speech because during the casual paired conversations, many of the participants discussed the topic of their accent and a number of participants talked specifically about how they pronounce /t/. One particular participant remarked that “we don’t pronounce our t’s around here”. In order to explore whether T-glottaling rates are linked to the participants’ sense of regional identity, their self-evaluations of identity (described in Chapter 4) were considered. Figure 6.17 presents the distribution of T-glottaling rates grouped according to whether the participants selected a national identity (*British* or *English*), a regional identity (*Yorkshire* or *West Yorkshire*), or a local identity (*Bradford*, *Huddersfield* or *Wakefield*).

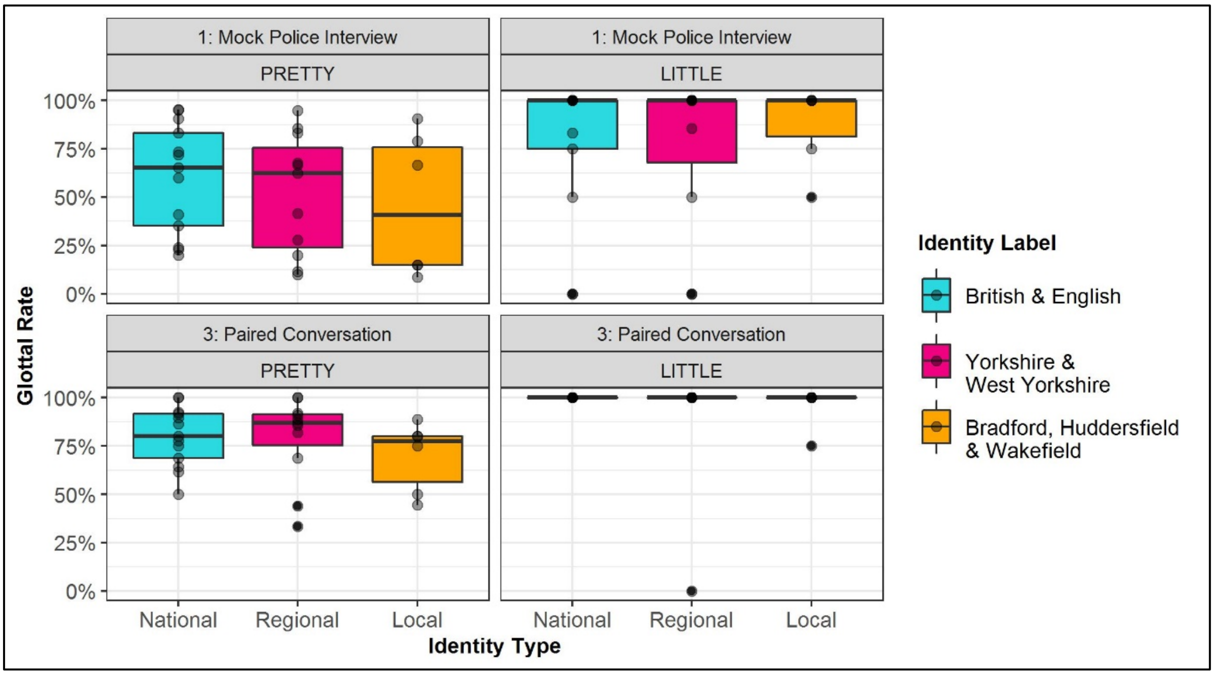


Figure 6.17. Distribution of T-glottaling rates across tasks and environments by identity type.

In Figure 6.17, it can be seen that the participants' T-glottaling rates in the LITTLE context were broadly similar regardless of which identity type the participants selected, in both the mock police interviews and the paired conversations. However, with regards to T-glottaling rates in the PRETTY context, there were some slight differences across groups, particularly in the paired conversation task. When considering the median values and the interquartile ranges visualised in Figure 6.17, it would appear that participants with a more local identity tend to have lower T-glottaling rates than those with a national or regional identity. This suggests that the participants' T-glottaling rates may be conditioned to some extent by their sense of identity, despite there being no local level regional differences overall. However, it must be recognised that there were relatively few data points across each identity type and more data would be required to examine this trend further.

6.6.2. Implications

6.6.2.1 Forensic speech science

This study provides a new set of population data for word-medial, intervocalic /t/ which generally shows that, in unstressed positions, West Yorkshire speakers tend to use both [ʔ] and [t], with a preference for the glottal variant. Furthermore, T-glottaling is blocked when /t/ appears in a foot-initial, stressed position. Overall, the finding that this specific phonetic variable is not regionally stratified at a very local level is good news from the perspective of a forensic expert involved in FSC casework. It would seem that the boroughs of Bradford, Kirklees and Wakefield could all be grouped together when describing how word-medial, intervocalic /t/ is typically realised in West Yorkshire. The implication of this finding is that when assessing the typicality of /t/ realisations within a particular individual from West Yorkshire, it may be appropriate to use the more broadly defined reference population of West Yorkshire, or even Yorkshire English, as opposed to defining the population more narrowly at the borough level. It is also potentially possible to go further and argue that in respect of this specific variable, younger male speakers of West Yorkshire English behave in much the same way as the majority of other Northern varieties of British English.

The findings in relation to differences across the separate phonetic environments considered in this study are also useful for forensic experts to be aware of when comparing tokens of intervocalic /t/ across different forensic samples. Specifically, this investigation confirms what much of the previous research relating to T-glottaling has shown in that T-glottaling is more likely to occur before a syllabic consonant than a vowel and it is extremely unlikely to occur when /t/ is in the stressed, foot-initial position. Care must always be taken to ensure that the expert is comparing like with like across forensic samples and therefore it is important that the phonetic environment is taken into account, in terms of word and syllable position as well as surrounding phonetic environment of /t/.

6.6.2.2 Sociophonetics

Regarding the implications of this study for the field of sociophonetics more generally, this investigation has provided an up-to-date account of how /t/ is realised in a region that has received relatively little attention. Findings of this investigation corroborate the general patterns that have been reported particularly in the North of England, whereby T-glottaling is on the increase, even in the word-medial intervocalic context, where it was traditionally regarded as being highly stigmatised. It is evident that the following phonetic context, syllabic position and prominence all affect how /t/ is realised and therefore it is suggested that these aspects should all be taken into account when conducting analyses of T-glottaling and variation in /t/ more generally.

In order to build on the findings of this study, it would be useful to investigate whether the patterns that have been observed in the WYRED dataset hold true in females from West Yorkshire and speakers from different age groups and social demographics. As T-glottaling has been shown to be correlated with speakers' social characteristics, including age, gender and to some extent social class it would be interesting to examine how much /t/ productions vary according to these factors across the region. Based on the findings of previous studies, it is expected that T-glottaling rates in particular may be higher within these young males than they might be in other cross-sections of the speech community.

6.7. Conclusion

This investigation has established how word-medial intervocalic /t/ is realised in a range of phonetic environments across three boroughs within West Yorkshire. Results show that speakers from Bradford, Kirklees and Wakefield broadly behave in the same way and do not index local level identity through use of this particular variable. Speakers from this area largely tend to alternate between using glottal variants and fully released voiceless alveolar plosives, with T-glottaling being the more prevalent form used overall. Using a generalised mixed effects linear regression analysis it was possible to determine that in this dataset, T-glottaling rates vary according to phonetic environment and speaking task. The following chapter will systematically explore how /t/ is realised across two tasks that vary in a range of stylistic ways and will consider to what extent the participants' productions are influenced by the effects of their respective interlocutors.

7. Phonetic accommodation in West Yorkshire /t/

7.1. Introduction

This chapter builds on the findings of the previous chapter in order to explore how the West Yorkshire participants' realisations of /t/ vary across different forensically-relevant speaking tasks that involve distinct speaking styles and interlocutors. An examination is presented of phonetic accommodation in word-medial, intervocalic /t/ by analysing productions from 30 West Yorkshire males from WYRED, and their respective interlocutors, partaking in a mock police interview scenario as well as a casual paired conversation. This analysis enables observations to be made about individual and group tendencies for the purpose of considering the extent to which /t/ productions are affected by accommodation (defined broadly here as the simultaneous influence of speaking style and interlocutor). These underpinning observations are subsequently evaluated in order to consider the extent to which accommodation could impact FSC casework.

This chapter is divided into 7 sections. This first section presents a description of how /t/ has been found to vary according to speaking style, and findings from previous speech accommodation studies involving /t/ are presented. In Section 7.2, the research questions and hypotheses for this investigation are set out and Section 7.3 provides a summary of the data that was used to analyse accommodation. Section 7.4 describes the methods used for this investigation and Section 7.5 provides the results in relation to each of the research questions posed in this chapter. Section 7.6 addresses the implications of these results from both a FSC casework perspective and also in terms of sociolinguistic research practices more generally, before the conclusions of this study are presented in Section 7.7.

7.1.1. Background research

In Chapter 6, it was established that the West Yorkshire participants in this study largely tend to alternate between use of [ʔ] and [t] for word-medial, intervocalic /t/, with glottal variants being the predominant form in an unstressed position. For this reason, the background research summarised in this section relates specifically to T-glottaling rather than any of the other /t/ variants that were discussed in the previous chapter. In Section 6.1.2, a range of social factors that have been found to correlate with use of T-glottaling were discussed, including social class, age and gender. In addition to social variation, T-glottaling has also been shown to vary according to style and audience design in previous studies.

In Stuart-Smith's (1999) investigation of T-glottaling in Glaswegian, a marked difference in T-glottaling rates between casual conversation and reading aloud a wordlist was observed, whereby far more glottal stops were produced in the casual conversations. This pattern was present in both the working class and middle class speakers, although it was noted that working class speakers tended to use [ʔ] when reading the wordlist more than the middle class speakers (1999b, p. 191). Evidence of style-shifting with respect to T-glottaling was also reported in Baranowski & Turton's (2015) investigation of variation in Manchester English. They found that T-glottaling was significantly more frequent in the interview style compared to when reading a wordlist, in both word-final and intervocalic contexts. In line with Stuart-Smith's findings, this trend was also observed for both working and middle class speakers, with working class speakers style-shifting to a slightly lesser degree by using [ʔ] more often than the middle-class speakers in the wordlist context. A number of other sociolinguistic studies that have examined variation in /t/ productions have reported higher rates of T-glottaling in casual styles compared to formal styles (Fabricius, 2002; Mees & Collins, 1999; Schleef, 2013). These findings are in line with the more general principal that the use of non-standard or low-prestige variants will typically increase in more informal speaking styles where less attention is paid to speech (Labov, 1972).

The fact that speakers style-shift in their rates of T-glottaling is perhaps unsurprising when considering how socially salient this feature is. It would seem that speakers of English have a high level of awareness of this linguistic feature, most likely in part as a result of the strong negative associations that were traditionally ascribed to this variable (particularly in relation to intervocalic T-glottaling) but also due to the extent to which it has spread across different parts of the UK. Evidence from non-native speakers of English also indicates that this is a salient feature of some varieties of British English. For example, Drummond (2011) investigated the acquisition of T-glottaling by native Polish speakers living in Manchester and found that the speakers who had been living in Manchester for the longest showed an increased rate of glottal replacement in pre-pausal and pre-vocalic word-final /t/. As previously mentioned, it has been suggested that accommodation tends to be more likely to occur in features that are socially salient (Cao, 2018; Smith & Holmes-Elliott, 2015; Trudgill, 1986), and therefore it seems highly likely that participants in the present investigation will adapt their glottal rates to some degree as a consequence of communicating with different interlocutors across different tasks.

The influence of audience design on T-glottaling rates was explored in Kirkham & Moore's (2016) analysis of two speeches by the former UK Labour Party leader, Ed Miliband, which were presented to audiences with different political views. By comparing Miliband's T-glottaling rates when speaking at the Trade Union Congress (TUC), with his T-glottaling rates when speaking at the Labour Party Conference (LPC), they observed that his use of [ʔ], particularly in *Britain* and *government,* seemed to reflect his alignment with the two different audiences. The word *Britain* was categorically pronounced with [ʔ] in the TUC speech and with [t] in the LPC speech. A similar pattern was observed in the word-final context in the word *government,* where 82% of tokens were realised with [ʔ] in the TUC speech and 62% were realised with [t] in the LPC speech. In the TUC speech, use of [ʔ] was said to occur in contexts where Miliband "express[es] alignment with his audience in terms of material action and thoughts/moral viewpoints" (2016, p. 105). In contrast to this, use of [t] in the LPC speech was observed when Miliband attempted to "establish credibility with his audience, without having to necessarily imply shared values" (2016, p. 105). Overall, these findings indicate that within-speaker variation in terms of T-glottaling rates can be used strategically to evoke certain social meanings, depending on the audience being addressed.

The influence of the interlocutor on T-glottaling rates was explored as part of Smith & Holmes-Elliott's (2017) investigation, in which speakers from Buckie interacted with a community insider and a community outsider. They initially predicted that speakers would use lower rates of the glottal variants when interacting with the community outsider, given the traditional stigmatisation of the feature (Smith & Holmes-Elliott, 2017, p. 16). However, no clear accommodation effects were apparent in the speech of the younger or middle-aged speakers and the opposite pattern was found within the older speakers. One possible explanation for this finding could be related to the interlocutors' respective glottal rates. The community outsider had higher glottal rates than the community insider and therefore the older speakers may have significantly increased their usage of [ʔ] in order to converge towards the higher glottal replacement user. The authors suggested that "as the rates of [ʔ] increase in the community, no such accommodation is warranted as speakers use [ʔ] as much or even more than the outsider" (2017, p. 26). This would provide an explanation for why the younger and middle-aged speakers did not accommodate, as their glottal rates were generally much higher than those of the older speakers.

Trudgill (1986, p. 8) reported that on examination of his own speech during interviews with his informants in his (1974) investigation of Norwich English, it became apparent that his glottal rates were positively correlated with those of his interviewees. Across the 10 interviews, Trudgill's glottal rates varied dramatically leading him to conclude that he must have subconsciously accommodated towards his informants. However, it is not known to what extent his informants were also accommodating towards him, nor can it be known how their glottal rates may have altered if Trudgill had maintained a more consistent glottal rate across all interviews. Although the specific purpose of Trudgill's original (1974) investigation was not to examine speech accommodation, his findings provide evidence of T-glottaling being susceptible to accommodation effects. This particular case study also highlights the need for researchers to be aware that, when conducting sociolinguistic interviews, their own speech has the potential to influence that of their participants. As discussed in Chapter 2, in the context of accommodation research the primary way to completely control the input of the interlocutor is to elicit speech via a non-interactive approach, such as a speech-shadowing task. However, the disadvantage of this method is that it is less ecologically valid than face-to-face human interaction.

7.2. Research questions and hypotheses

The study presented in this chapter aims to evaluate the extent to which speakers from West Yorkshire accommodate in their realisations of word-medial, intervocalic /t/, across two separate forensically-relevant scenarios. In order to do this, the following research questions are addressed:

1. What is the influence of the task on /t/ productions?
2. How do the participants' /t/ productions relate to those of their interlocutor?

a. Do any differences between tasks correlate with the usage of the interlocutors in the respective tasks?
b. Do participants produce more glottal variants when interacting with the interlocutor with higher T-glottaling rates?
c. Do participants' T-glottaling rates vary over the course of each paired task?

The first research question has been partially addressed in Chapter 6, where it was shown that the participants produced significantly more glottal variants in the casual paired conversations than in the mock police interview task, overall. In this chapter, further quantitative information is provided to explain the differences across tasks in detail. Comparisons are drawn between the full ranges of /t/ variants produced by the West Yorkshire participants, across three WYRED speaking tasks. Although it is acknowledged that there was insufficient data available to be able to draw any robust conclusions in relation to the answer message task, a brief summary of the /t/ data from this task is presented in the interest of transparency. Results are presented across tasks for /t/ tokens from each of the three phonetic environments considered in this investigation. It is anticipated that in addition to the differences already observed across tasks in glottal rates in the PRETTY and LITTLE environments, there may be further variation in minority variants and in the HOTEL environment.

In order to address the second research question, and its sub-parts, the distribution of /t/ variants produced by the participants' Task 1 and Task 3 interlocutors are first established and then compared with one another. It is predicted that the Task 3 interlocutors will largely have higher glottal rates than the Task 1 interlocutor, due to the differences in speaking style across tasks. The glottal rates of each participant and their respective interlocutors are then considered in relation to each other to determine how similar they are in this respect. In general, it is anticipated that participants' /t/ productions will correlate with their interlocutor and they will tend to have higher glottal rates when interacting with the interlocutor with higher T-glottaling rates, overall. It is also expected that participants whose interlocutor has high glottal rates will be more likely to have high glottal rates themselves. Because intervocalic T-glottaling is somewhat stigmatised, participants' T-glottaling rates in the formal mock police interview are expected to be similar to or lower than those of their interlocutor.

In explaining any differences in /t/ productions across tasks, it is not possible to separate the influence of the interlocutor from the speaking style. For this reason, it is also useful to consider how the participants' productions change over the course of each task. By comparing the first half of /t/ productions with the second half, it is possible to explore the effects of exposure to the interlocutor without the need to account for changes in speaking style. Based on the findings of Smith & Holmes-Elliott (2017), it is hypothesised that participants who generally have higher glottal rates may show less variation over the course of the task and be less affected by the influence of their interlocutors.

7.3. Data

For this investigation, the word-medial, intervocalic /t/ tokens produced by 30 WYRED participants during Tasks 1, 3 and 4 that were analysed in Chapter 6 were examined in relation to how they varied across speaking task. As mentioned in Chapter 6, there were very few tokens of intervocalic /t/ available in the Task 4 answer message recordings, with an average of only 2 tokens per participant. For this reason, it was not possible to use the participants' Task 4 realisations of /t/ as their baseline which could then be compared to their paired interactions in Task 1 and Task 3. Instead, this investigation focuses on how /t/ is realised in the mock police interview task (Task 1) and the casual paired conversation (Task 3) and considers how participants behave at the group level as well as individually. The Task 4 data is therefore only used to present preliminary visualisations of group trends across tasks and must be treated with caution. Table 7.1 provides a summary of the number of /t/ tokens included in this study, across the three tasks.

Table 7.1. Summary of /t/ tokens per task.

Task	Total	Average per participant	Median per participant
1	894	30	29
3	643	21	22
4	62	2	2

In addition to examining the participants' productions of word-medial, intervocalic /t/, it was also necessary to consider how each of their respective interlocutors in Task 1 and Task 3 realised this phonetic variable. As each participant was paired with another participant from the same borough in the Task 3 recordings, /t/ tokens were already analysed auditorily for both interlocutors in each pair. However, for the Task 1 recordings, participants spoke to a female researcher from Gateshead and therefore her /t/ tokens needed to be analysed separately. This was carried out using the same procedure as described in Section 6.4, whereby all suitable instances of word-medial, intervocalic /t/ were manually transcribed based on auditory perceptions. Again, Dr Erica Gold also transcribed 20% of these files. In total, 680 /t/ tokens were analysed across the 30 recorded conversations, with a median of 23 tokens per recording. Table 7.2 summarises the distribution of the researcher's word-medial, intervocalic /t/ tokens across the three phonetic environments considered in this study.

Table 7.2. /t/ tokens analysed in the researcher's 30 interviews, by phonetic environment.

Phonetic environment	**Number of tokens**
HOTEL	120
PRETTY	548
LITTLE	12
All environments	680

As can be seen in Table 7.2, the majority of the researcher's /t/ tokens were from the PRETTY environment and there were very few examples of LITTLE tokens. Table 7.3 provides a summary of the different realisations of word-medial, intervocalic /t/ that were observed within the researcher's speech. In line with the West Yorkshire participants, glottal variants produced by the researcher were coded as either [ʔ] or [ʔ]* depending on whether or not the glottal stop was visible in the acoustic signal, respectively. In both cases, the auditory perception was that /t/ had been replaced with an auditorily distinct glottal stop [ʔ]. Similar to the participants, the majority of /t/ tokens were realised as either glottal variants or fully released voiceless alveolar plosives. However, it appeared that the researcher used a higher proportion of [t] than the West Yorkshire participants, at least when considering all of the participants' tokens as a whole. Consequently, convergence towards the researcher in Task 1 would require an increased use of the voiceless alveolar plosive.

Table 7.3. Overall distribution of all /t/ variants for the researcher in Task 1.

Variant	[t]	[ʔ]*	[ʔ]	[ʔ͡ɾ]	[tˢ]	[ɾ]	[s]	[t͡s]	[ʔ͡t]
%	46.0	23.8	20.9	2.8	2.6	2.2	1.3	0.1	0.1
N	313	162	142	19	18	15	9	1	1

There were no instances of [ʔ] occurring in the HOTEL environment and therefore T-glottaling appears to be blocked in this context for the researcher, in the same way as it was for the participants. A range of the same minority variants that were observed in the participants' speech were also found in the researcher's recordings, however, there were no instances of [k] or of /t/ being fully elided. One additional variant was observed whereby /t/ was realised as a glottalised tap [ʔ͡ɾ]. Figure 7.1 provides an example of a token realised with this variant, in which there is a double articulation involving both an oral closure and a glottal closure. Similar to many of the glottal tokens, the auditory perception of the glottal articulation is created by a period of creaky voice at the end of the preceding vowel, which is then followed by an alveolar tap [ɾ]. Of the 19 tokens that were realised as [ʔ͡ɾ], 16 occurred in words where /t/ proceeded an /iː/ vowel, with 9 being from the word *amenities* and 5 from the word *forty*.

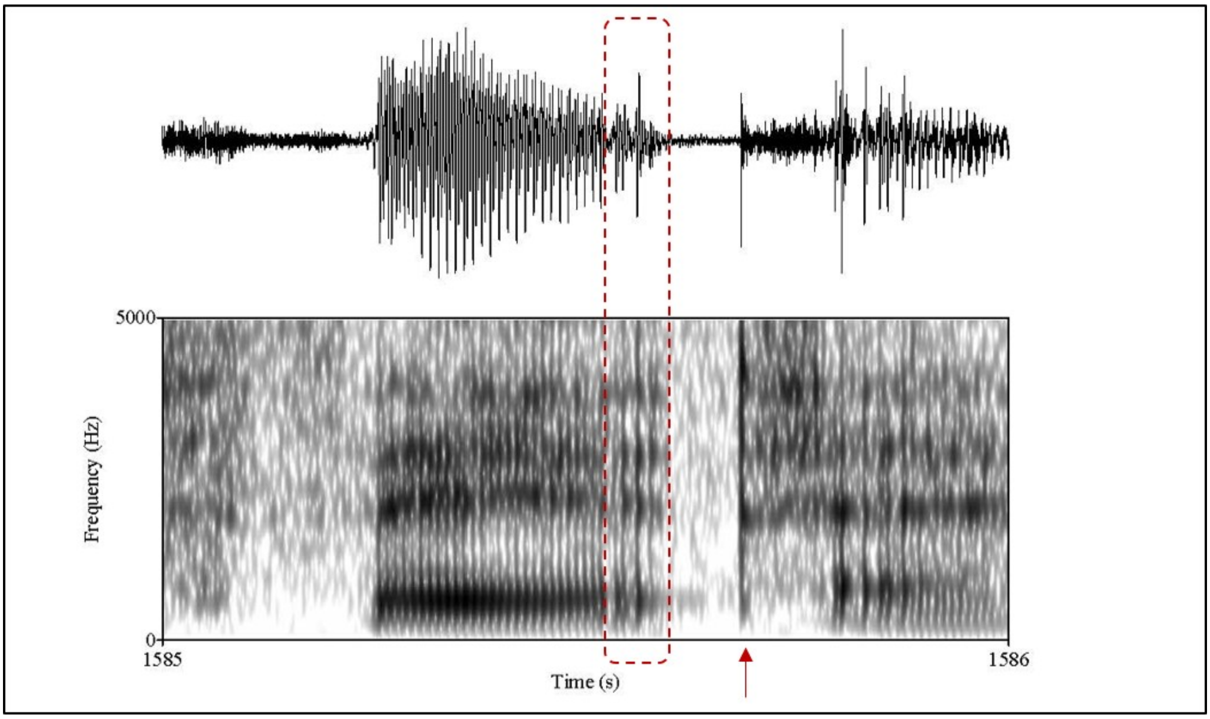


Figure 7.1. Spectrogram of *theatre* spoken by the researcher, showing [ʔ͡ɾ]. (The rectangle indicates the period of creaky voice and the arrow marks the [ɾ].)

7.4. Methodology

This section describes the methods used to examine phonetic accommodation in word-medial, intervocalic /t/ in West Yorkshire. This analysis involved comparing how realisations of /t/ varied across different forensically-relevant speaking tasks as well as considering how the participants' productions related to those of their respective interlocutors.

7.4.1. Measuring the influence of the task on /t/

Firstly, an assessment of the influence of task on /t/ productions was carried out by establishing whether /t/ was realised in different ways across different tasks. This assessment involved considering the overall distribution of /t/ variants within tasks as well as considering specifically whether T-glottaling rates varied according to task. Initial examinations of the influence of task on /t/ involved visualising the participants' /t/ data across all three tasks; however, due to the limited data available in Task 4, the main analysis involved only Task 1 and Task 3. In Chapter 6, a generalised mixed effects logistic regression analysis was presented which explored the effects of BOROUGH, TASK and PHONETIC ENVIRONMENT on T-glottaling rates, using the participants' Task 1 and Task 3 data. The full model that was used, with /t/ productions as the dependent variable, is shown again below:

T-glottaling.model = glmer(/t/ ~ BOROUGH + TASK + ENVIRONMENT + (1+TASK|PARTICIPANT), data=Data, family=binomial)

This chapter reiterates the findings of this analysis with respect to the effect of TASK and discusses whether there were any overwhelming patterns across the participants, whereby one task elicited significantly higher T-glottaling rates than the other. Graphs are presented to illustrate the findings of these quantitative analyses. The influence of PHONETIC ENVIRONMENT is also addressed and appropriately accounted for when examining how /t/ varied according to task. Regarding the effect of BOROUGH, it has been established that T-glottaling rates did not appear to vary significantly across the boroughs of Bradford, Kirklees and Wakefield and therefore the participants were not grouped by borough in the further analyses that follow in this chapter.

7.4.2. Measuring the influence of the interlocutor on /t/

In order to assess the potential influence of the interlocutor in each of the paired tasks, it was necessary to establish if /t/ productions were significantly different between the participants' interlocutor during the Task 1 interviews (the researcher) and during their Task 3 conversations (the West Yorkshire participants). To test this, a statistical analysis was carried out which examined the differences in T-glottaling rates across the two groups, whilst accounting for the effects of phonetic environment. This was similar to the generalised mixed effects logistic regression analysis described in Section 6.4.4 and was conducted using R version 3.5.0 (R Core Team, 2018) and lme4 version 1.1-21 (Bates, Mächler, Bolker, & Walker, 2015).

As with the West Yorkshire participants' data, all of the researcher's tokens involving glottal realisations of /t/ were coded as 1 and all non-glottal variants were coded as 0. As a fixed effect, GROUP was entered into the model, and this was treated as a categorical factor with two levels (Task 1 interlocutor and Task 3 interlocutor, with the former as the reference level). The fixed effect of PHONETIC ENVIRONMENT was also entered into the model with two categorical levels (PRETTY and LITTLE, with the former as the reference level). In contrast to the model described in Chapter 6, the fixed effect of TASK was not required as the Task 1 interlocutor's data all came from the Task 1 recordings and the Task 3 interlocutors' data all came from the Task 3 recordings. The fixed effect of BOROUGH was also left out of the model, as it had previously been established that there were no significant differences across boroughs and therefore this factor did not contribute to the model fit. As all of the individual interlocutors each produced multiple intervocalic /t/ tokens in each conversation, these responses could not be regarded as being independent from one another. To deal with this inter-dependency, a random effect for INTERLOCUTOR was entered into the model. The full model with /t/ productions as the dependent variable is presented below:

T-accommodation.model = glmer(/t/ ~ GROUP + ENVIRONMENT + (1|INTERLOCUTOR), data=Data, family=binomial)

In this model, the fixed effects of GROUP and PHONETIC ENVIRONMENT (without interaction terms) were used to predict T-glottaling rates. As random effects, there were intercepts for interlocutors. The model was fitted using maximum likelihood and to test the influence of each of the fixed effects on T-glottaling rates, p-values were obtained by likelihood ratio tests of the full model with all of the fixed effects against the model without each of the fixed effects in question.

In order to examine the extent to which the West Yorkshire participants' /t/ productions were influenced by accommodation effects, glottal rates were calculated for each participant across their Task 1 and Task 3 recordings, as well as those of their respective interlocutors in each task. Task 4 productions were not taken into account in this analysis as there was insufficient data available to be able to draw any reliable conclusions. This study focussed specifically on the amount the participants produced glottal variants because in Chapter 6 it was found that the variation in West Yorkshire is largely between [t] and [ʔ]. Two sets of glottal rates were calculated in R (R Core Team, 2018), firstly using all of the /t/ data from the PRETTY environment and secondly using the LITTLE tokens. HOTEL tokens were excluded as T-glottaling never occurred in this context. Although the T-glottaling rates in the LITTLE environment will be discussed briefly in this chapter, it should be noted that there were very few tokens available (median number of tokens per participant = 2 in both Task 1 and Task 3) compared to the PRETTY environment (median number of tokens = 21 and 16 in Tasks 1 and 3, respectively). Furthermore, the researcher only produced 12 LITTLE tokens across all 30 Task 1 conversations. For this reason, it was not possible to conduct a robust, quantitative analysis of individual accommodation behaviour using the LITTLE tokens.

In addition to considering how the participants' /t/ productions varied across tasks and how this related to the way in which their respective interlocutors realised /t/, it was also necessary to consider whether or not the participants' rates of T-glottaling appear to change within a task after spending more time interacting with an interlocutor. The effects of exposure to the interlocutor over time were assessed by dividing all of the /t/ data for each participant into two halves and comparing variants from the first half of /t/ tokens to those from the second half in each of the paired tasks. This comparison was performed using another generalised mixed effects logistic regression analysis, whereby the fixed effects of

LATENCY (early and late /t/ tokens), TASK (Task 1 and Task 3) and PHONETIC ENVIRONMENT (PRETTY and LITTLE) (without interaction terms) were used to predict T-glottaling rates. As random effects, there were intercepts for PARTICIPANTS as well as by-participant random slopes for the effect of TASK. The full model with /t/ productions as the dependent variable is presented below:

T-glottaling.latency.model = glmer(/t/ ~ LATENCY + TASK + ENVIRONMENT + (1+TASK|PARTICIPANT), data=Data, family=binomial)

7.5. Results

This section provides results in relation to each of the research questions under investigation in this study. Section 7.5.1 presents a descriptive analysis of how the West Yorkshire participants' /t/ productions varied across tasks and then goes on to summarise the results of the statistical analysis that was conducted. Section 7.5.2 focuses specifically on how the participants' productions relate to those of their respective interlocutors and provides details of accommodation behaviour both within and across tasks.

7.5.1. What is the influence of the task on /t/ productions?

The West Yorkshire participants examined in this investigation typically tend to alternate between using glottal variants and fully released voiceless alveolar plosives for word-medial, intervocalic /t/, with T-glottaling being the more prevalent form used overall. However, when comparing the distribution of variants across the three WYRED speaking tasks, it is evident that this phonetic variable is susceptible to change according to the task in which it is produced, with glottal variants being much more common in the casual paired conversation and the answer message tasks than in the mock police interview task.

7.5.1.1 Summary of distributions across tasks

Table 7.4 reports the distribution of different /t/ variants that were observed in the 30 West Yorkshire participants' speech across tasks. This data is also represented in the form of a bar chart in Figure 7.2. It should be noted that the West Yorkshire participants are not separated by borough in any of the figures that follow in this chapter, as it was determined in Chapter 6 that there were no significant differences across Bradford, Kirklees and Wakefield in terms of glottal rates and therefore this phonetic feature does not appear to be regionally stratified according to local borough. It is also important to reiterate here that the Task 4 data visualised in this section are only based on 62 tokens across all 30 participants and therefore apparent differences between Task 4, and Task 1 and 3 may be a consequence of this. For this reason, the Task 4 data has not been included in any of the statistical analyses in this chapter.

Table 7.4. Overall distribution of all /t/ variants across tasks.

Task		**[ʔ]***	**[t]**	**[ʔ]**	**Ø**	**[t^s]**	**[ɾ]**	**[t͡s]**	**[ʔ͡t]**	**[s]**	**[k]**
1	**%**	39.1	48.3	7.8	1.8	0.4	0.4	0.9	0.7	0.3	0.1
	N	350	432	70	16	4	4	8	6	3	1
3	**%**	52.4	22.4	16.0	4.8	3.6	0.8	-	-	-	-
	N	337	144	103	31	23	5	-	-	-	-
4	**%**	58.1	30.6	4.8	4.8	-	1.6	-	-	-	-
	N	36	19	3	3	-	1	-	-	-	-

It can be seen that in all three tasks the glottal variants [ʔ]* and [ʔ] together with the released alveolar plosive [t] account for over 90% of the data. When comparing the proportion of oral stops to glottal stops across tasks, it is clear that participants used a larger proportion of the standard variant [t] in Tasks 1 and 4 than in Task 3. In line with previous studies that have observed a trend whereby speakers increase their use of T-glottaling in less formal contexts, the participants in this study use this variant most often in the casual paired conversation and least in the formal, mock police interview task.

Table 7.4 and Figure 7.2 also show that there were differences in proportions of use of the other /t/ variants; with Ø, [t^s] and [ɾ] all occurring less often in Task 1 than Task 3, and [t͡s], [ʔ͡t], [s] and [k] only being present in Task 1. Overall, Ø, [t^s], [ɾ] only accounted for 3.1%, 1.7%, and 0.6% of the data respectively. There were no examples of [t^s] in the Task 4 recordings and only one example of [ɾ] in this task. The three most infrequent variants were only produced by a small subset of the participants in Task 1. Three participants realised /t/ as [t͡s], three used [ʔ͡t], two used [s] and only one produced /t/ as [k] on one occasion. Overall, as the minority variants only accounted for a relatively small proportion of the /t/ data, the analysis of accommodation in this study focusses primarily on the variation between [t] and [ʔ].

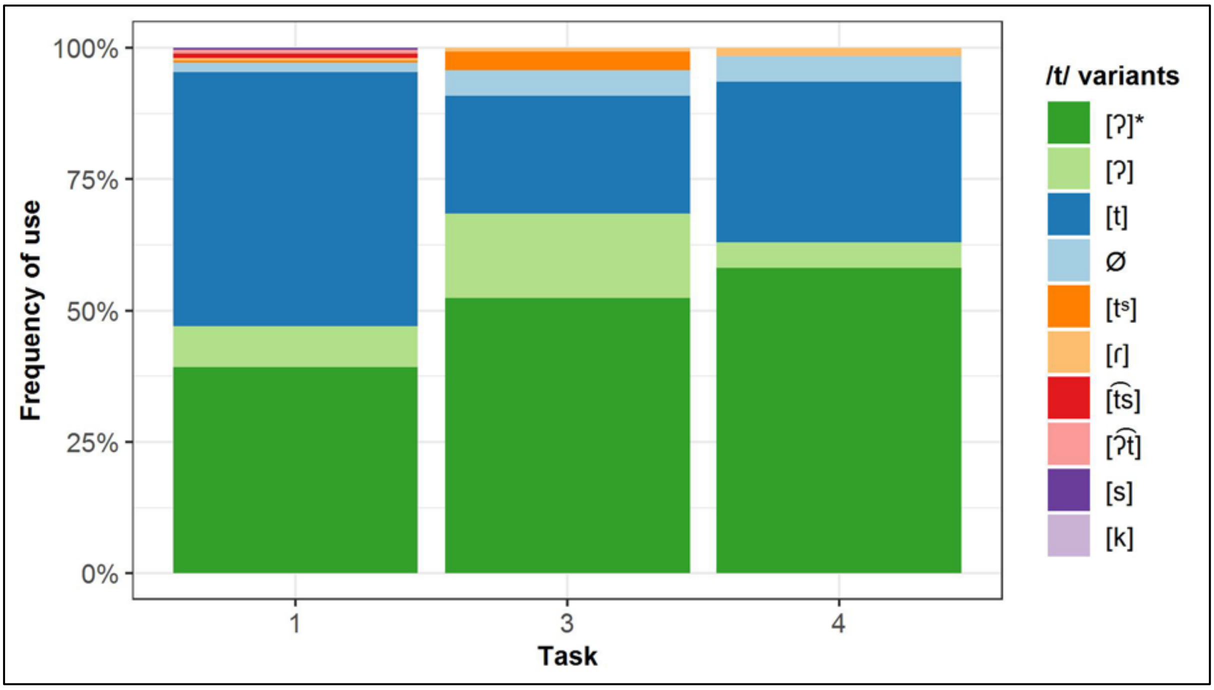


Figure 7.2. Distribution of the West Yorkshire participants' /t/ variants across tasks.

7.5.1.2 Distributions across tasks by phonetic environment

In Chapter 6, it was established that word-medial, intervocalic /t/ tokens can vary depending on the phonetic environment in which they occur and therefore it is necessary to account for phonetic environment effects when comparing distributions across tasks. For the purposes of this investigation, /t/ tokens were separated into three groups represented by the keywords HOTEL, PRETTY and LITTLE. Please see Section 6.4.3 for further details regarding each of these phonetic environments. In this West Yorkshire speech community, it is especially important to separate the HOTEL tokens from the PRETTY and LITTLE tokens when examining T-glottaling as it would appear that this phonetic feature cannot occur in the HOTEL environment and therefore a significantly different distribution of variants is expected. Figure 7.3 illustrates the distribution of /t/ variants across tasks, stratified by phonetic environment.

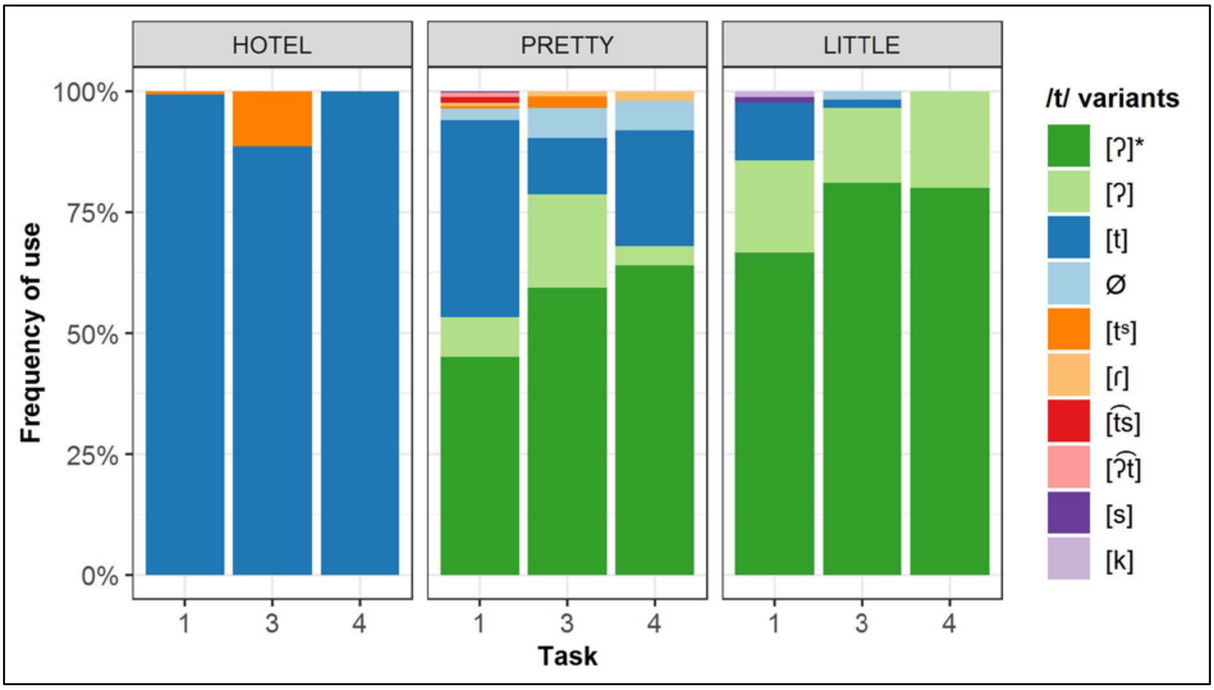


Figure 7.3. Distribution of the participants' /t/ variants by task and phonetic environment.

7.5.1.2.1 HOTEL

Beginning with the HOTEL environment, it can be seen that the variation is exclusively between [t] and [tˢ]. The affricated variant occurs most often in Task 3, with 11 /t/ tokens being realised as [tˢ] across the speech of seven particular individuals. This variant only occurs once in Task 1 and there are no examples of [tˢ] in the Task 4 data. Overall, based on the very low numbers of affricated variants present across Tasks 1 and 3, it does not appear that the participants' /t/ productions in the HOTEL environment are influenced by the effects of speaking style or the interlocutor.

7.5.1.2.2 PRETTY

In the PRETTY environment, the most common variants are [ʔ]*, [ʔ] and [t] with the glottal variants far outweighing [t] across all three tasks. The largest proportion of [t] tokens occur in Task 1, the most formal of the three tasks, and [t] is least frequent in the casual paired conversation. There are also subtle differences in the proportions of minority variants used across Tasks 1, 3 and 4 in this environment. For example, in line with the HOTEL environment, there were more examples of [tˢ] in Task 3 than in Task 1. Full elision of /t/ was also more common in Task 3 than in Task 1, and all instances of [t͡s], [ʔ͡t] and [s] occurred in Task 1, although there were relatively few examples of each.

7.5.1.2.3 LITTLE

In the LITTLE environment, the vast majority of tokens were realised as [ʔ] or [ʔ]*. The largest proportion of [t] variants occurred in Task 1, with the only token of [t] in Task 3 and none in Task 4. There were only three tokens that were not realised as either [ʔ]*, [ʔ] or [t] and these were realised as [s] and [k] in Task 1 and Ø in Task 3.

7.5.1.3 Glottal rates across tasks by phonetic environment

Overall glottal rates were lower in Task 1 than Task 3, and comparatively lower in the PRETTY environment than the LITTLE environment. The boxplots presented in Figure 7.4 visualise the participants' individual glottal rates across tasks in these two environments. Individual glottal rates are not presented for the Task 4 data as there were too few tokens available per participant for these values to be meaningful. It must also be acknowledged that the number of tokens analysed across the three phonetic environment varies greatly. Across all 30 West Yorkshire participants there were 1190 PRETTY tokens, 262 HOTEL tokens and 147 LITTLE tokens. The average number of LITTLE tokens per participant is less than 5, and therefore the findings relating to individual T-glottaling rates in the LITTLE environment are only tentative. Nevertheless, there appears to be a larger range in glottal rates across the participants in Task 1 than in Task 3, in both phonetic environments.

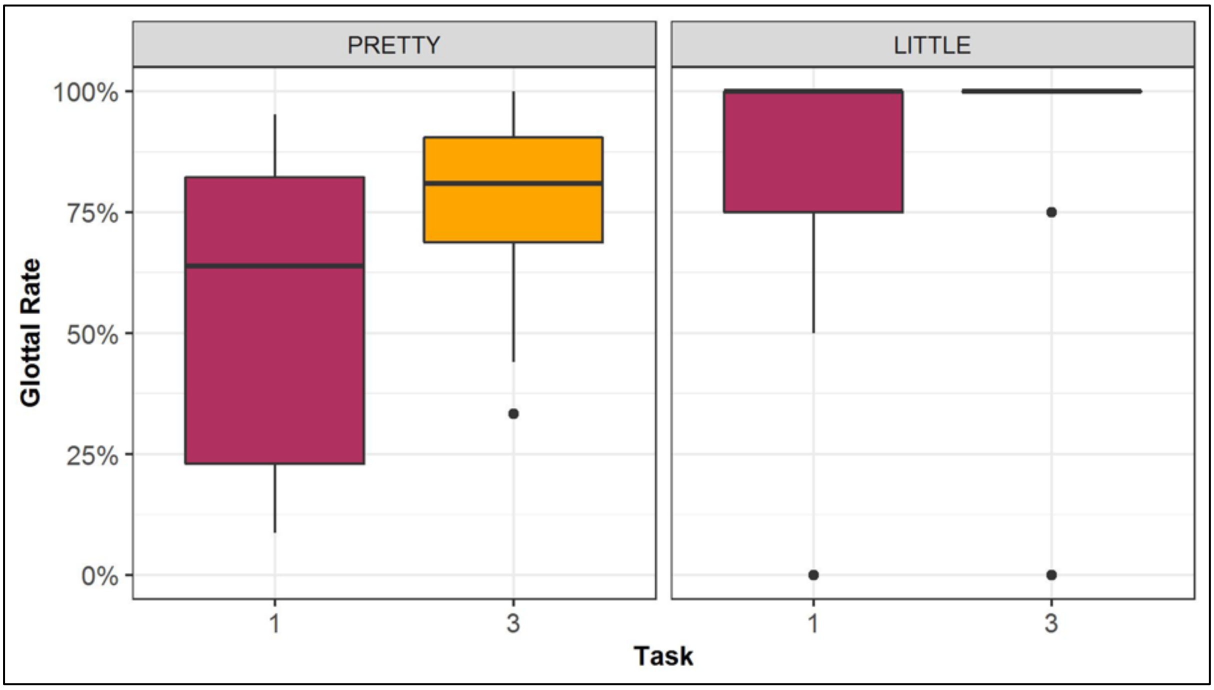


Figure 7.4. Glottal rates of the West Yorkshire participants by task and phonetic environment.

The generalised mixed effects logistic regression analysis presented in Section 6.4.4 examined the relationship between T-glottaling rates and the fixed effects of BOROUGH, TASK and PHONETIC ENVIRONMENT, using all of the PRETTY and LITTLE /t/ data from Tasks 1 and 3. The p-values that were obtained by likelihood ratio tests of the full model with the fixed effect of TASK against the model without the effect of TASK showed that there were significant differences in terms of T-glottaling rates across Task 1 and Task 3 (χ^2 (1) = 15.128, p <0.0001), with the glottal variant being more frequent in the casual paired conversation (Task 3) than in the mock police interview (Task 1). Additionally, the phonetic environment had a statistically significant effect on T-glottaling (χ2 (1) = 58.795, p <0.0001), with glottal variants being most frequent in the LITTLE environment. BOROUGH did not have a significant influence on T-glottaling rates, as no clear differences were observed between /t/ productions of the participants from Bradford, Kirklees and Wakefield.

The fact that T-glottaling rates were significantly higher in Task 3 than in Task 1 indicates that the differences across the two tasks may have influenced how word-medial, intervocalic /t/ is realised. In relation to the differing speaking styles, the direction of change across tasks appears to be in line with the findings of previous studies of T-glottaling which have shown that this non-standard variant is more frequently observed in casual contexts as opposed to formal speaking styles. In addition to contrasting levels of formality, it is also necessary to consider the influence of the interlocutor as a supplementary, or possibly an alternative, explanation for the differences in T-glottaling rates across the two tasks. Although it is not possible to separate the effects of speaking style from those of the interlocutor, in this investigation accommodation is defined broadly to refer to adaptations in speech across situations involving different interlocutors and speaking styles, with the aim being to establish the extent to which /t/ productions vary across forensically-relevant speaking tasks.

7.5.2. How do the participants' /t/ productions relate to those of their interlocutor?

This section examines how the participants' /t/ realisations varied across the mock police interview task and the casual paired conversation, with an emphasis on how they may have been influenced by the participants' respective interlocutors in each task. Accordingly, this section begins by setting out how /t/ was realised by the female researcher from Gateshead, during the 30 Task 1 interviews. This is then compared to how the West Yorkshire participants realised /t/ during the Task 3 conversations, where they each interacted with another participant from the same borough as themselves. It is important to emphasise here that while all participants were interviewed by the same person in Task 1, in Task 3 each participant spoke to a different person. It may therefore be expected that higher levels of variation will be present across the interlocutor data for Task 3 than for Task 1.

7.5.2.1 Task 1 and Task 3 interlocutor comparison

The distribution of /t/ variants within the interlocutor's speech during the Task 1 mock police interviews and the interlocutors in the Task 3 paired conversations are presented in Figure 7.5. It would appear that in respect of the HOTEL tokens, overall the Task 1 and Task 3 interlocutors were relatively similar in that they mostly varied between [t] and [t^{s}] with the affricated variant being very infrequent. Of the 120 HOTEL tokens that the interlocutor in Task 1 produced, there were only two examples of /t/ being realised as [t^{s}] and one example of [t͡s]. None of these minority variants occurred when interacting with a participant who produced these variants. In comparison, there were 11 instances of [t^{s}] occurring in the speech of seven of the interlocutors in Task 3 but in almost all cases these tokens were produced when interacting with a participant who did not use this variant.

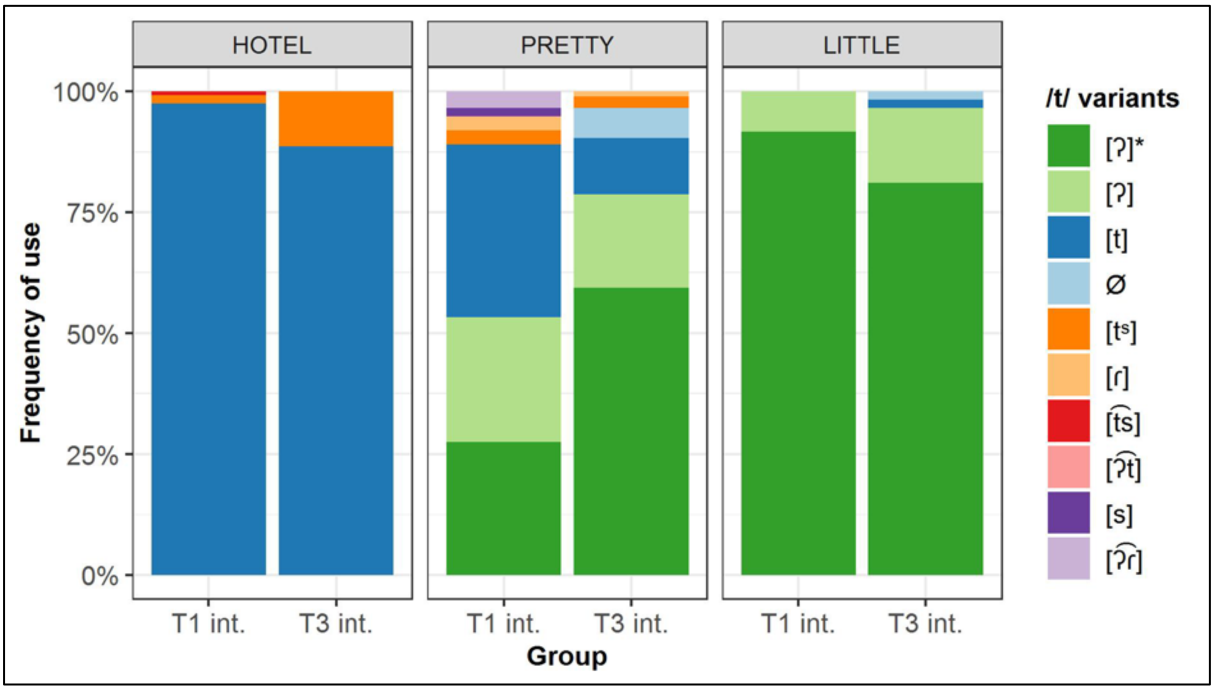


Figure 7.5. All /t/ variants for the Task 1 and Task 3 interlocutors, by phonetic environment.

The Task 1 and Task 3 interlocutors were also very similar in how they realised /t/ in the LITTLE environment. All of the LITTLE tokens that the Task 1 interlocutor produced were realised with glottal variants, however, it is important to reiterate that there were only 12 LITTLE tokens produced across all 30 Task 1 conversations. For the interlocutors in Task 3, all except two of the 58 tokens that were produced in this environment were realised as a glottal variant. There was only one instance of /t/ being realised as [t] and one instance of elision. Based on the comparisons drawn between the Task 1 and Task 3 interlocutors' productions in the HOTEL and LITTLE environments, it would seem that the participants tended to be exposed to similar variants when interacting with each of their respective interlocutors across the two tasks. However, the distributions of /t/ variants for tokens in the PRETTY environment appear to reflect more variation across the interlocutors.

In the PRETTY environment, approximately 90% of the Task 1 and Task 3 interlocutors' tokens were realised as either [ʔ]*, [ʔ] or [t]. However, the Task 1 interlocutor used a higher proportion of [t] than the Task 3 interlocutors, overall. In terms of minority variants, the Task 1 interlocutor also realised /t/ as [ʔ͡ɾ], [tˢ], [ɾ], and [s]. Across the speech of the Task 3 interlocutors, there were examples of /t/ being realised as [tˢ] and [ɾ], as well as instances where /t/ was fully elided. Now if we focus specifically on T-glottaling rates, Figure 7.6 shows that rates were generally higher within the Task 3 interlocutors than the Task 1 interlocutor. The boxplot also illustrates that the T-glottaling rates in the PRETTY environment were more spread across the different Task 3 interlocutors compared to the interlocutor in Task 1 across the 30 interviews. This is unsurprising given that the participants' interlocutor remained consistent in Task 1 whereas every participant interacted with a different speaker in the Task 3 recordings.

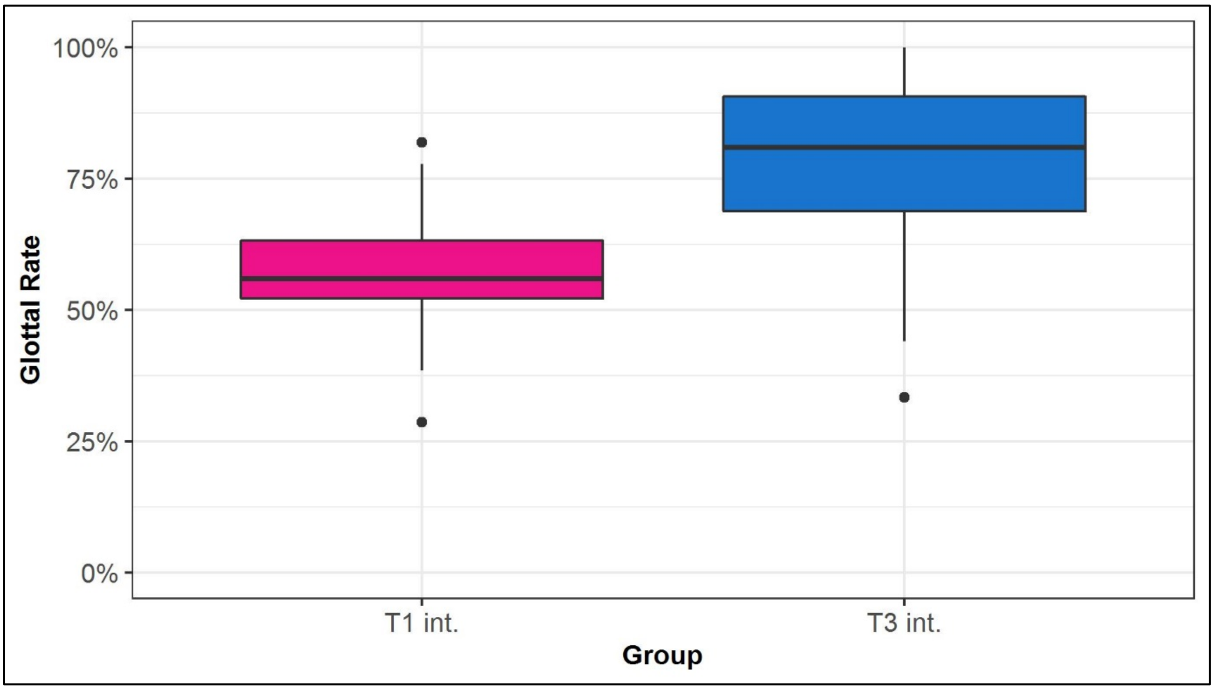


Figure 7.6. Glottal rates in PRETTY /t/ tokens from the Task 1 and 3 interlocutors' conversations.

The generalised mixed effects logistic regression analysis described in Section 7.4.2 examined the relationship between T-glottaling rates and the fixed effects of GROUP and PHONETIC ENVIRONMENT in order to compare the t-glottaling rates of the interlocutors across the two tasks. Table 7.5 shows the results of the model of best fit, as determined by a generalised mixed effects logistic regression analysis, based on 1106 observations of word medial, intervocalic /t/. Any positive estimates in the regression coefficients indicated more use of T-glottaling within that category, when compared with the reference level intercept. Therefore, the results show that the Task 3 interlocutors were more likely to use a glottal variant than the Task 1 interlocutor. However, the p-values that were obtained by likelihood ratio tests of the full model with the fixed effect of GROUP against the model without the effect of GROUP showed that there were no significant differences in T-glottaling rates across the Task 1 and Task 3 interlocutors. This finding appears to be due to higher levels of variability in the Task 3 conversations arising from there being different speakers across conversations, as opposed to one consistent interlocutor in the Task 1 interviews. When re-running the statistical model with the random effects adjusted so as to treat all Task 3 data as though it was from the same interlocutor (as it is in Task 1), the differences in glottal rates of PRETTY and LITTLE tokens between groups were found to be significant. However, it was necessary to account for the fact that the Task 3 data was produced by multiple speakers, and therefore the original model was retained.

Table 7.5. Coefficients of a mixed-effects logistic regression model of T-glottaling , with random intercepts for interlocutor (SD = 0.87).

	Estimate	**Std. Error**	**z-value**	**Pr (>|z|)**
Intercept (Task 1 interlocutor, PRETTY)	0.284	0.874	0.325	0.745
Task 3 interlocutor	1.164	0.898	1.297	0.195
LITTLE	2.712	0.729	3.719	<0.0001

In order to assess the influence of the phonetic environment on T-glottaling, p-values were also obtained by likelihood ratio tests of the full model against the model without the fixed effect of PHONETIC ENVIRONMENT. In line with the previous statistical analyses presented in Chapter 6, this test confirmed that the phonetic environment had a statistically significant effect on T-glottaling ($\chi2$ (1) = 29.807, $p < 0.0001$), with glottal variants being most frequent in the LITTLE environment.

Although the T-glottaling rates of the Task 1 and Task 3 interlocutors were not significantly different, there does appear to be a tendency for the Task 3 interlocutors to have higher T-glottaling rates than the Task 1 interlocutor in the PRETTY environment. For this reason, at least some of the participants may be more likely to have higher T-glottaling rates in Task 3 than in Task 1. However, it is difficult to predict how the participants may accommodate as a consequence of the influence of their interlocutor, as it is possible that the participants could converge, diverge or not accommodate at all. Furthermore, it is necessary to consider that it is not just the interlocutor that varies across the two tasks but also the nature of the task, the formality and power dynamics between the interlocutor pairs. As this combination of factors are often at play in forensically-relevant scenarios, it is useful to examine how they can contribute to within-speaker variation in T-glottaling rates.

7.5.2.2 Accommodation in T-glottaling rates across tasks

While it is not possible to discuss the full range of /t/ variants for all 30 participants in detail here, Appendix 6 provides individual bar charts to visualise the distribution of /t/ variants for each participant and their respective interlocutor in Tasks 1 and 3, across phonetic environments. This section focuses specifically on how the participants' T-glottaling rates varied across Tasks 1 and 3 and considers these findings in relation to the T-glottaling rates of their respective interlocutors. Table 7.6 provides a summary of the average and median glottal rates that were observed in the participants' PRETTY and LITTLE /t/ tokens across tasks. Overall, there was a marked increase in both the average and median glottal rates from Task 1 to Task 3, in the PRETTY and LITTLE environments. T-glottaling rates were also higher for LITTLE tokens than for PRETTY tokens, with median glottal rates of 100% for LITTLE tokens both Tasks.

Table 7.6. Participants' average and median glottal rates in PRETTY and LITTLE /t/ tokens.

Phonetic environment	Task	Average	Median
PRETTY	Task 1	54.1%	63.9%
	Task 3	77.6%	80.9%
LITTLE	Task 1	78.9%	100%
	Task 3	94.3%	100%

Table 7.7 reports the average and median glottal rates for the Task 1 interlocutor across the 30 interviews. There is no equivalent table for the Task 3 interlocutor, as each participant interacted with another participant during this task and therefore the Task 3 summary data

in Table 7.6 also relates to the interlocutors. By comparing the Task 1 interlocutor's values with those of the West Yorkshire participants it can be seen that on average the participants' glottal rates for PRETTY tokens during Task 1 were relatively similar to the interlocutor's rates. Both the participants and the Task 1 interlocutor had a median glottal rate of 100% in the LITTLE tokens, although the average glottal rate was lower for the participants than the interlocutor due to all of the interlocutor's LITTLE token being realised as a glottal variant.

Table 7.7. Task 1 interlocutor's average and median glottal rates in PRETTY and LITTLE /t/ tokens.

Phonetic environment	Average	Median
PRETTY	56.5%	55.9%
LITTLE	100%	100%

As mentioned previously, there were insufficient tokens for a detailed analysis of accommodation in the LITTLE environment and therefore the following results just focus on the PRETTY environment. Table 7.8 summarises the changes in PRETTY T-glottaling rates across tasks by presenting the number of participants who increased their glottal rates from Task 1 to Task 3, the number of participants who decreased from Task 1 to Task 3 and the number of participants who had consistent rates across tasks. T-glottaling rates were considered to be consistent if the change from Task 1 to Task 3 was less than 5%. Summary statistics are also provided regarding the amount of change from Task 1 to Task 3. Based on previous literature cited in Section 7.1.1 which has shown that T-glottaling is more common in more casual speaking styles compared to formal styles, we might expect to find that all participants would have higher glottal rates in the Task 3 casual paired conversation compared to the more formal, mock police interview in Task 1. However, what we see is that although most participants had higher glottal rates in Task 3 than Task 1; there were some participants who displayed the opposite pattern and some who remained consistent across tasks. This finding could indicate that some form of accommodation has occurred in respect of T-glottaling rates.

Table 7.8. Changes in glottal rates from Task 1 to Task 3 in PRETTY /t/ tokens.

Δ	No. of participants	Average Δ	Median Δ	Minimum Δ	Maximum Δ
Increase	21	35.8%	27.8%	5.8%	90%
Same	6	-	-	-	-
Decrease	3	16.8%	15.2%	5.9%	29.2%

The values in Table 7.8 highlight that the participants whose T-glottaling rate decreased from Task 1 to Task 3 typically changed to a lesser degree than those who increased from Task 1 to Task 3. In order to determine whether or not the participants' T-glottaling rates in Task 1 were correlated with their T-glottaling rates in Task 3, a Spearman's rank correlation test was conducted. The results show that there was a moderate positive correlation between the T-glottaling rates across the two tasks ($r_s = 0.41$, $n = 30$, $p<0.05$). This suggests that high glottal rates in one task are somewhat likely to lead to high rates in the other and similarly low rates in one task may lead to low rates in the other.

7.5.2.2.1 Individual T-glottaling rates

Figure 7.7 displays the individual T-glottaling rates across Task 1 and Task 3 for each of the 30 West Yorkshire participants. The Task 1 data is shown in maroon and the Task 3 data is shown in orange. Participants are ordered according to the size of the difference in T-glottaling rates between tasks, with the largest decrease from Task 1 to Task 3 on the left and the largest increase from Task 1 to Task 3 on the right.

It can be observed that the Task 1 glottal rates are lower than the Task 3 glottal rates for the majority of participants. However, the T-glottaling rates of participants #019, #049 and #033 decreased from Task 1 to Task 3, with participant #019 displaying the largest decrease. These participants appear to be deviating from the expected style patterns which could suggest that they are being influenced by their respective interlocutors' /t/ productions. Six participants (#046, #020, #067, #036, #064 and #006) had consistent glottal rates across Tasks 1 and 3.

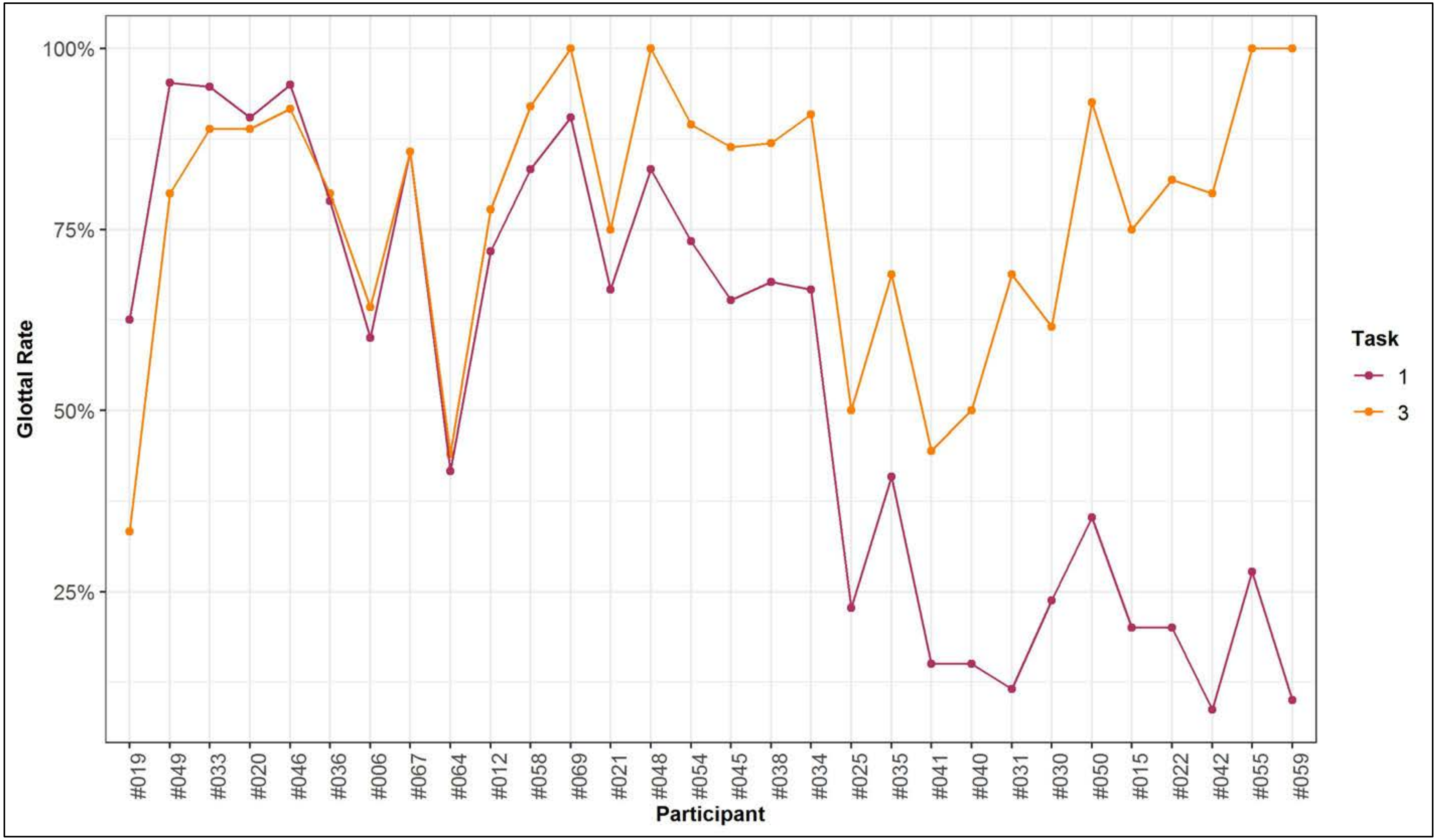


Figure 7.7. Glottal rates in PRETTY /t/ tokens across participants during Task 1 and Task 3.

It is interesting to note that the glottal rates for this group of participants ranged from approximately 40% up to 95%, as can be seen in Figure 7.7. For this reason, it cannot be said that participants who have particularly high or low T-glottaling rates were more likely to be consistent across tasks involving different interlocutors and speaking styles, and vice-versa. It could be the case that these individuals are just less inclined to style-shift in general or perhaps T-glottaling in word-medial, intervocalic /t/ is not a feature that they would consider to be strongly associated with a particular speaking style. Alternatively, these participants could be maintaining similar glottal rates across tasks in order to accommodate towards (or away from) their interlocutor.

In Figure 7.7, all of the participants to the right of participant #006 on the x-axis have higher glottal rates in Task 3 than in Task 1, with the amount of change ranging from 5.8% up to an increase of 90%. It can also be seen that some of the participants with the lowest glottal rates in Task 1 displayed the largest increase in glottal rates from Task 1 to Task 3 (e.g. #042 and #059). While this trend is in line with previous findings relating to the influence of speaking style on T-glottaling, it is also possible that the differences in the participants' /t/ productions across tasks may have been influenced by their respective interlocutors' productions.

7.5.2.2.2 Influence of interlocutor on T-glottaling rates

To determine whether or not the participants' T-glottaling rates were correlated with the T-glottaling rates of their respective interlocutors in each of the tasks, a Spearman's rank correlation test was conducted, firstly for Task 1 and then for Task 3. For each of the tasks, the results show that there was no statistically significant correlation between the T-glottaling rates of the participants and their interlocutors. This finding indicates that it is not possible to predict a participant's T-glottaling rates based solely on their interlocutor's rates. In order to further explore any potential effects of accommodation on T-glottaling, Figure 7.8 presents the glottal rates of each participant across Task 1 and Task 3 together with the glottal rates of their respective interlocutors in each task. In this graph, participant rates are represented by circles and solid lines, whereas interlocutor rates are represented by triangles and dashed lines. Task 1 data is shown in maroon and Task 3 data is in orange. Participants are ordered by glottal rates in Task 1, with the highest on the left and the lowest on the right.

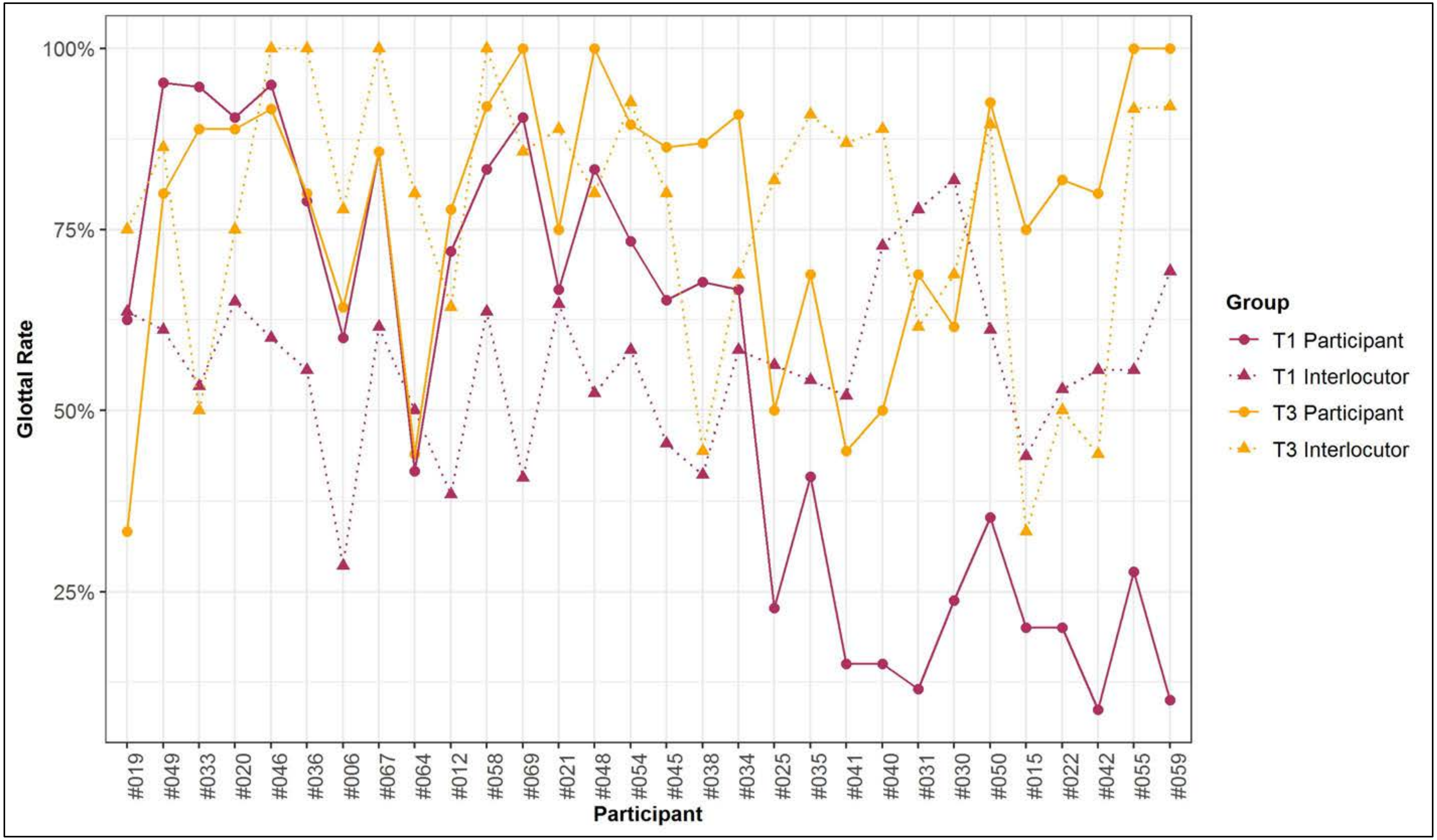


Figure 7.8. Glottal rates in PRETTY /t/ tokens of participants and their respective interlocutors across Task 1 and Task 3.

Of the three participants who had lower glottal rates in Task 3 than in Task 1, two participants (#019 and #049) interacted with an interlocutor in Task 3 who had higher glottal rates than their interlocutor in Task 1. For participant #019, the data in Figure 7.8 would suggest that he converged towards his Task 1 interlocutor but diverged away from his Task 3 interlocutor, based solely on differences in T-glottaling rates between interlocutor pairs. Participant #049, on the other hand, appeared to diverge from his interlocutor in Task 1 with a much higher glottal rate than his interlocutor whereas in Task 3 his glottal rate was more similar to his interlocutor but slightly lower. Regarding participant #033, it can be seen in Figure 7.8 that the differences between his T-glottaling rates across tasks was relatively small, with only a slight increase in Task 1 compared to Task 3. The glottal rates of participant #033's two interlocutors were considered to be consistent as there was only a 3% difference between them. Overall, participant #033's glottal rates were much higher than those of his interlocutors across both tasks and therefore it could be the case that he was diverging away from his interlocutors in both tasks, doing so to a slightly lesser extent in Task 3.

Regarding the participants who maintained consistent glottal rates across Task 1 and 3, it is evident that their respective interlocutors in Task 3 all had higher glottal rates than their interlocutor in Task 1. However, the amount by which the glottal rates differed between each of the interlocutor pairs varied substantially. For example, participant #046 has a glottal rate of approximately 90% in Task 1 and Task 3, which is similar to the glottal rate of his Task 3 interlocutor (100%) but much higher than that of his Task 1 interlocutor (60%). While the difference in formality levels across tasks might be expected to cause the participant to produce fewer glottal realisations in the mock police interview than in the casual paired conversation, it is possible that this participant is diverging away from the police interviewer in the Task 1 conversation by maintaining a high proportion of glottal productions. Participant #064 displays the opposite trend whereby he has relatively low glottal rates across Task 1 and 3 (approximately 40%), which is only slightly lower than the glottal rate of his Task 1 interlocutor (50%) but much lower than that of his Task 3 interlocutor (80%). In this instance, it would appear that the participant is diverging away from his Task 3 interlocutor and converging towards his interlocutor in Task 1.

Of the participants whose glottal rates increased from Task 1 to Task 3, the majority interacted with someone who had higher glottal rates in Task 3 (15 out of 21 participants) and two interacted with interlocutors across the two tasks who had consistent glottal rates. In these particular instances it is difficult to determine whether participants were more heavily influenced by the context of the interaction in terms of formality and speaking style or the glottal rates of their interlocutor. However, there were four participants (#30, #015, #031 and #042) for whom the speaking style appears to have had a greater effect on their /t/ productions, as their glottal rates were higher in Task 3 despite their respective interlocutor in this task having lower glottal rates than their interlocutor in Task 1. Furthermore, all of these participants displayed relatively large increases in T-glottaling rates from Task 1 to Task 3, with an average increase of 56.7%, ranging from 37.7% to 71.3%. All four of these participants had much lower glottal rates than their respective interlocutors during Task 1 (46.8% lower on average) and three out of four had higher glottal rates than their respective interlocutors during Task 3.

7.5.2.3 Accommodation in T-glottaling rates within tasks

In addition to considering how the participants' T-glottaling rates compared with each of their respective interlocutors across Task 1 and Task 3, this investigation also considered whether or not the participants' rates of T-glottaling appear to change within a task after spending more time interacting with an interlocutor. The effects of exposure to the interlocutor over time were assessed by dividing all of the /t/ data for each participant into two halves and comparing variants from the first half of /t/ tokens to those from the second half, in each of the paired tasks. This was tested using another statistical analysis, described fully in Section 7.4.2, whereby the fixed effects of LATENCY, TASK and PHONETIC ENVIRONMENT (without interaction terms) were used to predict T-glottaling rates. Table 7.9 shows the results of the model of best fit, as determined by a generalised mixed effects logistic regression analysis, based on 1256 observations of word medial, intervocalic /t/. P-values were obtained separately by way of model comparison using likelihood ratio tests.

As expected and in line with the previous analysis presented in Chapter 6, the influence of the speaking task on T-glottaling was found to be statistically significant ($\chi 2$ (1) = 14.579, p

<0.0001), with glottal variants being most frequent in Task 3. Additionally, the PHONETIC ENVIRONMENT was also revealed to have a statistically significant effect on T-glottaling (χ2 (1) = 58.584, p <0.0001), with glottal variants being most frequent in the LITTLE environment.

Table 7.9. Coefficients of mixed-effects logistic regression model of T-glottaling , with random intercepts for participant (SD = 1.506) and by-participant random slopes for task (SD = 1.140).

	Estimate	Std. Error	z-value	Pr (>|z|)
Intercept (Early, Task 1, PRETTY)	0.037	0.301	0.124	0.901
Late	0.418	0.146	2.859	0.004
Task 3	1.160	0.266	4.368	<0.0001
LITTLE	2.243	0.350	6.409	<0.0001

To test the influence of exposure to the interlocutor on T-glottaling, p-values were obtained by likelihood ratio tests of the full model, against the model without the fixed effect of LATENCY. This analysis revealed that LATENCY had a statistically significant effect on T-glottaling (χ^2 (1) = 8.068, p <0.005), with the glottal variant being more frequent in the late tokens than the early tokens, overall. This finding indicates that the participants may have adapted their word-medial, intervocalic /t/ productions over the course of the task as a consequence of increased exposure to their interlocutor, leading to an increase in T-glottaling. In order to explore this finding further, glottal rates were calculated for each participant and their respective interlocutor across each set of early and late PRETTY /t/ tokens and absolute differences between each interlocutor pair's glottal rates were established. The absolute differences from late tokens were subsequently subtracted from the absolute differences from the early tokens in order to calculate the amount of change in interlocutor differences in T-glottaling rates over the course of the conversation.

For clarity, an example of the data used to examine how participant #033's glottal rates compared with his interlocutor's rates over the course of Task 3 is provided in Table 7.10. In this example, we can see that the participant's glottal rate for the late tokens was slightly higher than the rate for the early tokens and the opposite was true for his interlocutor. The absolute difference between the glottal rates of participant #033 and his interlocutor was larger for the late tokens than for the early tokens, suggesting that the participant diverged away from the interlocutor over the course of the task.

Table 7.10. Example of glottal rates for early and late PRETTY /t/ tokens of participant #033 and his Task 3 interlocutor and the differences between them.

	Participant glottal rates	**T3 Interlocutor glottal rates**	**Differences**	**Early → Late difference ∆**
Early	84%	54%	30%	Larger +17%
Late	92%	45%	47%	

It should be noted that only tokens from the PRETTY environment were included in this analysis, as there was limited data available in the LITTLE environment. However, caution must still be taken when interpreting these findings because the PRETTY /t/ data was divided into two halves for this analysis which meant that, for some participants, there were very few tokens in each set (median tokens per participant, per half = 10 for Task 1 and 7.5 for Task 3). For this reason, the results of this section need to be considered alongside the broader analysis of accommodation across tasks presented in Section 7.5.2.

Table 7.11 presents a summary of the number of participants whose glottal rates either increased, decreased or remained consistent across the early and late /t/ tokens, respectively. Again, T-glottaling rates were considered to be consistent if the difference between rates from early and late tokens was less than 5%. For those participants whose glottal rates either increased or decreased across latency conditions, the average amount by which they changed is also presented. Overall, the figures in Table 7.11 indicate that the differences between glottal rates from the early and late tokens were generally quite mixed, with slightly more participants increasing with more exposure to their interlocutor than those decreasing or remaining the same. There were four participants whose glottal rates increased from early to late tokens in both tasks and five participants whose glottal rates decreased from early to late tokens in both tasks. In cases where the participants' glottal rates did increase or decrease within a task, the rates changed by approximately 25% on average.

Table 7.11. Changes in glottal rates from early to late PRETTY /t/ tokens in Tasks 1 and Task 3.

		Early to late token glottal rate ∆	
Task	∆	**Number of participants**	**Average ∆**
1	**Increase**	17	20%
	Same	6	-
	Decrease	7	22%
3	**Increase**	14	24%
	Same	6	-
	Decrease	10	25%

It had been hypothesised that participants who generally have higher glottal rates may show less variation over the course of the task and be less affected by the influence of their interlocutors. In order to test this hypothesis, a Spearman's rank correlation test was conducted to determine whether or not the participants' overall T-glottaling rates in a task were correlated with the differences in T-glottaling rates from the early to late tokens in that task. The results show that there was no statistically significant correlation for either the Task 1 or Task 3 data. This finding suggests that the amount by which a participant's glottal rates altered from the first half to the second half of /t/ productions within a task was not directly related to their overall T-glottaling rate within that task.

By calculating the differences in glottal rates between interlocutor pairs, across early and late tokens separately, it was possible to explore how the participants' glottal rates varied in relation to the interlocutor over the course of the task. In cases where there was a decrease of more than 5% in the difference between an interlocutor pair's glottal rates from early to late tokens, this was taken to indicate convergence towards the interlocutor. When the difference between an interlocutor pair's glottal rates increased by more than 5%, this suggested divergence. Table 7.12 sets out the number of participants who were considered to have converged, diverged and stayed the same over the course of each task, as well as the average amount by which the distance between the interlocutor pair's glottal rates differed. In both Task 1 and Task 3, approximately half of the participants converged in terms of their T-glottaling rate and the average amount by which they converged was slightly higher in Task

1 than Task 3. While there were four participants who neither converged nor diverged in Task 1, there were just two who did this in Task 3. There were more participants who diverged in Task 3 than in Task 1 and the average amount by which they diverged was also larger in Task 3 than in Task 1.

Table 7.12. Early to late PRETTY /t/ token changes for T-glottaling rate differences between interlocutor pairs (across Task 1 and Task 3).

		∆ in early to late glottal rate differences	
Task	∆	**Number of participants**	**Average ∆**
1	**Increase (divergence)**	11	23%
	Same (maintenance)	4	-
	Decrease (convergence)	15	26%
3	**Increase (divergence)**	14	33%
	Same (maintenance)	2	-
	Decrease (convergence)	14	19%

The results reported in Table 7.12 could be interpreted to indicate that increased exposure to the interlocutor led to changes in the T-glottaling rates of the majority of participants (24/30). While the influence of the interlocutor and speaking style are conflated when comparing T-glottaling rates across Tasks 1 and 3, the speaking style is considered to be relatively consistent across the production of the early and late tokens within a task (although it is possible that participants may have become slightly more relaxed as they settled into each task). Overall, these results suggest that there was a fairly even split between participants who converged and participants who diverged within each paired task.

7.6. Discussion

This section begins by summarising the findings of the present investigation and then briefly discusses how these findings relate to those of previous studies. The implications of the findings for FSC casework, and for researchers in the fields of forensic speech science and sociophonetics more generally are also highlighted.

7.6.1. Summary of findings

As one of the fundamental aims of this study was to examine phonetic accommodation within forensically-relevant scenarios, speech from multiple speaking situations involving different interlocutors and speaking styles was analysed. As it is not possible to separate the effects of the interlocutor from the influence of speaking style, when comparing /t/ productions across tasks, a broad definition of accommodation is applied in this study to capture the simultaneous influence of both of these factors. In keeping with this, the findings of this study indicate that many participants do accommodate across the mock police interview task and the casual paired conversation. Specific findings in relation to each of the research questions (RQs) addressed in this investigation are summarised below.

RQ1: What is the influence of the task on /t/ productions?

Analysis of the /t/ data across the three WYRED speaking tasks revealed that [ʔ], [ʔ]* and [t] were the most prevalent variants, across all three tasks. When stratified by phonetic environment, it was determined that there were some subtle differences in proportions of minority /t/ variants across tasks, however, there were too few examples to be able to draw reliable conclusions about the influence of task on their distribution. There was also limited data available from the answer message task to be able to consider this task in the formal analysis. For this reason, comparisons of /t/ productions between the mock police interview task and the casual paired conversation were conducted and differences were observed within each of the three linguistic environments.

In the HOTEL environment, /t/ was mainly realised as [t], with one token realised as [t^s] in Task 1 and 11 in Task 3. Due to the low numbers of affricated tokens across tasks, there was insufficient evidence to suggest that productions of /t/ in this phonetic environment are influenced by effects of speaking style or interlocutor. In the PRETTY and LITTLE environments, there were significant differences in T-glottaling rates across tasks, with a much higher proportion of [t] in the mock police interviews than in the casual conversations. These findings were in line with the predictions in Section 7.2, as well as the findings of other studies that have reported increases in T-glottaling rates in more casual speaking styles. Overall, the average amount by which the participants' T-glottaling rates differed across tasks was 27%, however, there were high levels of variation between speakers, with T-glottaling rate differences ranging from no change to a difference of 90% across tasks.

RQ2: How do the participants' /t/ productions relate to those of their interlocutor?

a. **Do any differences between tasks correlate with the usage of the interlocutors in the respective tasks?**
b. **Do participants produce more glottal variants when interacting with the interlocutor with higher T-glottaling rates?**
c. **Do participants' T-glottaling rates vary over the course of each paired task?**

It was predicted that the interlocutor in the mock police interview would generally tend to have lower T-glottaling rates than the interlocutors in the casual paired conversation. It was also predicted that the participants' glottal rates would be higher in Task 3 than in Task 1, in line with their respective interlocutors. Analysis of the T-glottaling rates within the interlocutors' speech showed that the Task 3 interlocutors tended to have higher glottal rates than the Task 1 interlocutor. However, results of a generalised mixed effects regression analysis showed that the differences were not statistically significant. A Spearman's rank correlation test also found that the participants' T-glottaling rates were not correlated with those of their respective interlocutors in either Task 1 or Task 3. This finding indicated that there was no straightforward relationship between the participants and their interlocutors, at the group level. At the individual level, it was found that although most participants had higher glottal rates in Task 3 than in Task 1, the opposite was true for three participants and six were consistent across tasks. The amount by which the participants' glottal rates varied across task was also highly variable.

It was also hypothesised that participants would tend to have higher T-glottaling rates when they spoke to their interlocutor with the higher T-glottaling rates. This was found to be true for the majority of participants but not for all of them. Of the 24 participants whose T-glottaling rates changed across tasks, 17 had higher glottal rates when interacting with the interlocutor with higher glottal rates (compared to their interlocutor in the other task), and for 15 of these participants T-glottaling rates were highest in Task 3. It was also predicted that participants with higher T-glottaling rates may show less variation within a task and be less affected by the influence of their interlocutors. However, no correlation was found between the participants' glottal rates across a whole task and the amount by which their early and late glottal rates changed.

The results regarding accommodation behaviour over the course of the individual tasks are in some respects similar to those regarding accommodation across tasks. In both cases, there were significant differences in T-glottaling rates at the group level which were corroborated by statistical analyses. However, analysis of the individual results revealed that there was plenty of between-speaker variation in terms of both the degree and direction of the changes across and within tasks. At the group level, an examination of the variation in glottal rates within tasks showed that T-glottaling rates were higher in the second half of /t/ tokens than in the first half, and examination of the variation across tasks showed that T-glottaling rates were higher in Task 3 than in Task 1. In respect of these findings, it was observed that the average increase from Task 1 to Task 3 was higher than the average increase from early to late tokens in Task 1 (35.8% versus 20%) and in Task 3 (35.8% versus 24%). These findings suggest that T-glottaling rates were generally more heavily affected by changes across tasks than within them, as we might expect. Nevertheless, it must be emphasised that this was not the case for all participants.

7.6.2. Comparison with previous studies

To put the findings of the present study into context, the T-glottaling rates of the West Yorkshire participants in this study can be compared with those of speakers from Buckie (Smith & Holmes-Elliott, 2017) and Manchester (Baranowski & Turton, 2015).

As described in Section 7.1.1, in Smith & Holmes-Elliott's (2017) investigation of T-glottaling in Buckie, comparisons were drawn between the rates of speakers interacting with a community insider and a community outsider. In both cases, speech was elicited using a Labovian-style sociolinguistic interview technique. In relation to the interlocutors included in the present study, the community insider would be similar to the Task 3 interlocutor whereas the community outsider would be similar to the Task 1 interlocutor. However, it is potentially relevant to highlight that in Smith & Holmes-Elliott's study the community outsider had higher glottal rates than the insider whereas the opposite was true in the present study. In terms of social demographics, the young speakers from Buckie were considered to most closely match the participants in the present study. Overall, no clear accommodation effects were apparent in the speech of the younger Buckie speakers. In contrast to this finding, fairly large accommodation effects were evident in the present study. One reason for these larger differences could be related to the fact that the tasks in the present investigation varied not only by interlocutor but also in terms of the formality and nature of the activity. While the mock police interview involved the participants answering a series of questions and having to hide any incriminating information, in a potentially stressful situation, the casual paired conversations were much more relaxed and involved the participants simply talking about any topics that they wanted. In terms of overall T-glottaling rates, the younger Buckie speakers had average rates in the vicinity of 85-95% in the phonetic environments equivalent to the PRETTY and LITTLE environments in this study. Interestingly, for this group of speakers T-glottaling rates were slightly higher in the PRETTY environment than the LITTLE environment, which is opposite to the trend observed in the present study where the equivalent average T-glottaling rates were 65% and 90%, respectively.

Baranowski & Turton (2015) presented style-shifting rates in intervocalic T-glottaling in formal speech elicited via a wordlist task and from casual speech elicited using a sociolinguistic interview framework (where speech was coded by the researcher as either 'narrative' or 'careful' speech). In relation to the speaking styles included in the present study, the speech spoken in narrative mode from the sociolinguistic interview is considered to be closest to the Task 3 casual paired conversation whereas the more careful speech is perhaps somewhere in between Task 1 and Task 3 in terms of formality. In terms of social demographics, the middle class male speakers from Manchester were considered to most closely match the West Yorkshire participants in the present study. Across all Manchester speakers, the narratives and the careful speech showed no significant difference in T-glottaling rates and were therefore grouped together as 'casual speech'. The middle class males had an average glottal rate of 52% in the casual context dropping to 3% in the wordlist task. In the equivalent phonetic environment for the present study (PRETTY), the average T-glottaling rates of the participants dropped from 77.6% in the casual paired conversation to 54.1% in the more formal mock police interview task. Although there was no equivalent task in the present study that could be compared to Baranowski & Turton's wordlist task, it would appear that the West Yorkshire participants had slightly higher glottal rates overall when comparing their speech to the spontaneous speech of the Manchester speakers.

Focusing solely on the spontaneous speech, it can be seen that the West Yorkshire participants' T-glottaling rates varied across tasks to a much greater extent than the Manchester speakers' rates varied across the narrative and careful speaking styles. One reason for this could be that the WYRED tasks varied in terms of formality and interlocutor, whereas the Manchester speakers communicated with the same interlocutor throughout the sociolinguistic interview and therefore only the speech style varied. Although it is possible that the style differences were more extreme in the present study, this finding could also provide further evidence to suggest that the interlocutor affected the West Yorkshire participants' T-glottaling rates.

Compared to the findings of the two studies described here, it would appear that the West Yorkshire participants are broadly similar to the Manchester and Buckie speakers in terms of the average T-glottaling rates observed. However, the West Yorkshire participants appear to accommodate to a much stronger degree across two semi-spontaneous speaking styles, than the participants in the other studies. It is possible that this is due to the combined influences of different interlocutors and speaking styles present in this investigation. Overall, examination of word-medial, intervocalic /t/ at the individual level revealed that the effects of speech accommodation cannot be straightforwardly predicted based on the speaking task alone. Although there were clear trends at the group level, often there were participants whose /t/ productions, and specifically T-glottaling rates, deviated from these patterns.

7.6.3. Implications

7.6.3.1 Forensic speech science

From the perspective of a researcher in the field of forensic speech science, understanding the extent to which T-glottaling rates can vary within an individual is useful for the purposes of carrying out FSC casework. Using the data in this investigation, it is not possible to explicitly determine whether the interlocutor or speaking style has the greatest influence on T-glottaling. Nevertheless, applying a broad definition of accommodation to capture the combined influence of both of these factors, it is evident that for some participants quite extreme differences in T-glottaling rates are observed across tasks. In Figure 7.7, the glottal rates of the participants plotted towards each end of the x-axis can be seen to vary dramatically across tasks, with participants displaying low glottal rates in Task 1 and high rates in Task 3 showing the most marked differences. For instance, participant #059 has a glottal rate of only 10% in the mock police interview task versus a rate of 100% in the casual paired conversation.

In FSC cases requiring the comparison of speech samples involving different speaking styles (e.g. a formal police interview versus a casual conversation with a friend), an expert would take this disparity into account when evaluating any similarities and differences in their findings. When interpreting the findings, experts can consult previous studies which have documented within-speaker variability in a range of different scenarios. With regards to speaking style differences, the literature would generally point towards lower use of non-standard or less prestigious variants in more formal speaking situations. In the case of T-glottaling in particular, it has been shown that higher rates would typically be expected in more casual speaking styles. For this reason, the cases where participants' T-glottaling rates do not conform to the expected patterns are of most interest, and potential concern, from the perspective of FSC casework. For example, participant #019 has a glottal rate of 62.5% in the more formal context but this drops to 33.3% in the casual speaking style. A difference in this direction and of this magnitude could be interpreted as evidence of the two samples being produced by different speakers. Although auditory analysis of /t/ is just one of the many analyses that would typically contribute to a FSC case, it is possible that accommodation in one feature has the potential to lead to accommodation in others. Therefore, an awareness of possible within-speaker variation that could arise as a result of accommodation is useful for interpreting evidence in FSC casework.

7.6.3.2 Sociophonetics

The findings presented in this chapter and in Chapter 6 provide an up-to-date account of how word-medial, intervocalic /t/ is realised across the West Yorkshire boroughs of Bradford, Kirklees and Wakefield. While the previous chapter demonstrated that in terms of glottal rates, this phonetic feature does not appear to be regionally stratified according to local borough, this chapter has provided an insight into how glottal rates can vary within an individual across different speaking situations. The results of this investigation have shown that, within this group of West Yorkshire speakers at least, T-glottaling rates tend to increase in more casual speaking situations compared with more formal, interview-based tasks. However, it is important to recognise that this was not the case for all individuals. Furthermore, it is worth noting that the average T-glottaling rates are still relatively high, even in the most formal task considered in this study (54.1% for PRETTY and 78.9% for LITTLE).

It would appear that T-glottaling rates in the word-medial, intervocalic context can vary drastically across tasks involving different speech styles and interlocutors. One reason for this could be due to T-glottaling being most highly stigmatised and socially salient in this context, and it is likely that accommodation in T-glottaling rates would be less extreme in word-final /t/, for example. However, this was not formally examined in the present study. Nevertheless, in terms of practical implications for sociophoneticians, the findings of this study highlight the benefits of eliciting speech samples from a range of different speaking situations in order to explore within-speaker variation. In addition to providing information about how /t/ productions typically vary across different speaking styles, the results of this investigation also provide an indication of how much individual variation in T-glottaling rates can occur, even within a fairly homogenous population of speakers. The results of this investigation contribute to the existing body of literature relating to within- and between-speaker variation in T-glottaling and /t/ variation more broadly.

7.7. Conclusion

This chapter has built on the findings of the previous chapter (which investigated how /t/ was produced across the boroughs of Bradford, Kirklees and Wakefield) by exploring how T-glottaling rates vary across tasks involving different interlocutors and speaking styles. By considering how T-glottaling rates varied both within and across tasks, as well as how the participants’ rates related to those of their respective interlocutors, it was possible to evaluate how susceptible this phonetic feature is to the influence of speech accommodation.

The findings of this study illustrated that the way in which the West Yorkshire participants realise word-medial, intervocalic /t/ can be affected by the speaking task and the influence of the interlocutor. This is evident from the fact that participants’ T-glottaling rates changed not only across tasks but also within tasks, which could indicate that increased exposure to their interlocutor influenced their /t/ productions. The fact that T-glottaling is an extremely socially salient feature that speakers are typically aware of in their own speech and in the speech of others was thought to be one of the main reasons why high levels of within-speaker variation were present. In terms of the practical implications of this finding for FSC casework, it is advised that accommodation is something that experts need to be aware of when carrying out an analysis, as this study has shown that some speakers are more heavily influenced by the effects of speech accommodation than others.

8. Discussion and fictional FSC example

This chapter provides an overview of the findings of each of the main research questions addressed in this thesis. Section 8.1 discusses the findings in relation to regional variation across the West Yorkshire boroughs of Bradford, Kirklees and Wakefield. Section 8.2 considers the influence of speech accommodation on the West Yorkshire FACE vowel and intervocalic /t/. Section 8.2 also presents further results of a case study of one participant who was found to accommodate to a high degree. This is presented in the form of a fictional FSC case example and a consideration of the influence of accommodation on its outcome is provided. Section 8.3 subsequently goes on to highlight the main implications of the findings of this research for the field of forensic speech science and some recommendations and considerations for FSC casework and LADO are outlined. In Section 8.4, the limitations of the research presented in this thesis are acknowledged and Section 8.5 sets out ideas for further work. Section 8.6 concludes the chapter.

8.1. Regional variation in West Yorkshire

RQ1: How much local level variation exists across the West Yorkshire boroughs of Bradford, Kirklees and Wakefield with respect to the FACE vowel and intervocalic /t/?

In Chapters 4 and 6, levels of variation across participants from three separate boroughs were evaluated with respect to the FACE vowel and productions of word-medial, intervocalic /t/, respectively. The findings of these studies demonstrated that these two speech parameters are not regionally stratified in the same way. In the case of the FACE vowel, there were significant acoustic differences in F2 values across the three boroughs; with participants from Wakefield tending to have the most front realisations and participants from Kirklees having the least front productions. In addition to these acoustic differences, impressionistically, realisations of FACE also varied slightly according to borough. However, when considering all tokens from all participants, a range of variants including [ɛ], [eː], [ɪ], [eɪ], [ɛɪ] and [æɪ] were observed within each of the three boroughs. In contrast to the FACE findings, there was no evidence of local level regional variation in intervocalic /t/ productions across the West Yorkshire boroughs. Across all three boroughs, in unstressed positions intervocalic /t/ was mainly found to be realised as [ʔ] or as [t], to a lesser degree. T-glottaling also appeared to be blocked in all boroughs when /t/ occurs in a stressed position.

It was interesting to find that there were differing levels of regional variation between FACE and intervocalic /t/ as this may in part explain the range of opinion across the participants as to how much the West Yorkshire accent varies by borough. As part of the Task 3 paired conversations, participants were provided with a set of prompt cards including one which encouraged discussion of the topic of the West Yorkshire accent(s) (see Figure 8.1). Whilst discussing this topic, three pairs of participants included in the present study stated that they could not perceive any difference between the accents of Bradford, Kirklees and Wakefield. Conversely, five pairs claimed that there were clear differences between at least some of the accents across the West Yorkshire boroughs. Many other WYRED participants who were not included in the present investigation also talked of being able to perceive accent differences not only between the separate boroughs of West Yorkshire but also between individual areas within boroughs. Interestingly, the participants who said there were no differences between the accents of Bradford, Kirklees and Wakefield all selected more general identity descriptions when they were asked to self-evaluate their regional identity as part of the data collection process for WYRED (e.g. British, English or Yorkshire as opposed to a specific borough). It may therefore be the case that people are less likely to have a strong local identity at the borough level if they are unaware of, or do not consider there to be, distinguishing accent features between boroughs.

The meta-commentary from the Task 3 recordings relating to local level variability did not explicitly reference either T-glottaling or FACE realisations and generally tended to be quite vague. For instance, participant #025 from Bradford claimed that, in relation to the Bradford accent, Kirklees is “lighter” but still Northern while Bradford is “more broad and more deep”. Additionally, participant #045 from Wakefield stated that Wakefield “does not have an accent” whereas Leeds is “really broad” and “quite plain”. Participant #038 from Wakefield also claimed that Wakefield is relatively neutral and less regional, whereas Leeds has a “stronger accent” and he provided an example of the GOAT vowel in the word *go* being realised as [ɔː] in Leeds. Elaborating on this, participant #038 also stated that Wakefield and Kirklees were less distinct from one another whereas Bradford was considered to be "more blunt" because these speakers "don't drag their words out as much".

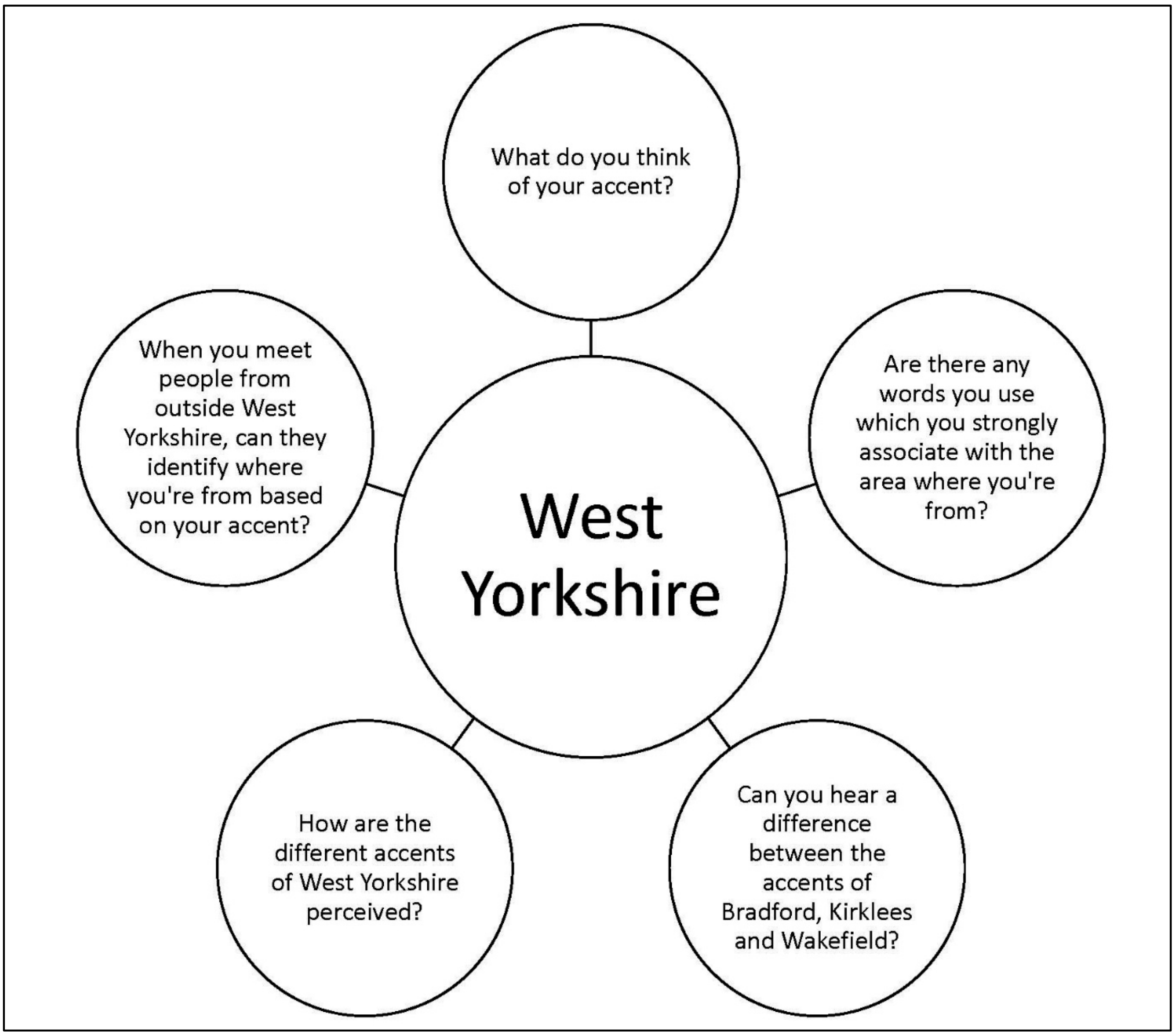


Figure 8.1. Prompt card about West Yorkshire presented to participants during Task 3.

Much of the commentary also seemed to relate to voice quality features. For example, participant #064 from Bradford stated that Huddersfield (Kirklees) is "harsher" and more "deep" compared to Bradford. Whereas, participant #046 from Kirklees claimed that the accents of Wakefield and Leeds are "harsher" and "more common" than Kirklees with Leeds sounding "tinny". In a study by Gold, Kirchhübel, Earnshaw, & Ross (forthcoming), macro and micro-regional variation of voice quality was explored in West Yorkshire English and SSBE, using a different subset of 60 WYRED participants and 20 DyViS participants respectively. The results of this study revealed that although there were some slight differences in voice quality profiles between the West Yorkshire and SSBE groups, there was no evidence of local level voice quality differences between the boroughs of Bradford, Kirklees and Wakefield.

It was suggested by a participant from Kirklees (#055) that one explanation for local level differences in accents could be due to the fact that there is a larger population of migrants in Bradford than there are in Kirklees and Wakefield. This claim can be supported by the census data reported in Section 3.1 which indicated that Bradford was the most diverse of the three boroughs in terms of ethnicity and language use whereas Wakefield appeared to be the most homogeneous borough. In addition to differences between boroughs, one participant (#033) also mentioned that the West Yorkshire accent varies across different generations, with older Yorkshire speakers having a more "broad" accent than younger speakers. It must be taken into account that the participants included in this study were all young, male speakers with English as their first and only language and therefore some of the perceived differences across the three areas may not have been fully represented in this investigation.

RQ2: Should reference population data for West Yorkshire speakers be separated by metropolitan borough or is it appropriate to group the boroughs more broadly as "West Yorkshire English"?

An important area of research in the field of forensic speech science involves examining the effects of specificity in reference population selection for FSC evidence. Hughes & Foulkes (2017) explored this topic by examining the influence of regional background specificity on FSC likelihood ratio outputs. In this study, likelihood ratios were computed using development and reference data which were regionally matched (SSBE) and mixed (general British English) relative to the test data. By evaluating the distributions of $\log_{10}$ likelihood ratios and the performance of the two systems, it was established that the strength of different speaker evidence was considerably greater and validity was consistently better when using the closely matched system compared with the mixed system. Additionally, it was observed that the impact of using more closely defined reference populations varied depending on the variable being investigated. It was said to be more crucial that an appropriate reference population is selected when examining regionally stratified speech parameters than it would be if the feature does not vary strongly across different regions. It was concluded that "while the more specific population produced better validity and stronger evidence, there is a greater associated risk of incorrectly defining the population." (Hughes & Foulkes, 2017, p. 3775).

The findings of Hughes & Foulkes (2017) reinforce the importance of researching phonetic variation at the local level in order to better understand the extent to which different phonetic parameters are regionally stratified. Although the present investigation did not calculate likelihood ratios to explicitly test the influence of delimiting the reference population at the regional level of West Yorkshire versus the local borough level, the results of this investigation are still considered relevant for this area of research. The findings in relation to regional variation in FACE, presented in Chapter 4, show that F2 is a marker of local level regional identity in West Yorkshire. Consequently, typicality assessments in relation to FACE F2 values should be carried out using reference population data which is narrowly defined at the borough level, in order to avoid under- or overestimations of the strength of evidence. However, as FACE F1 and F3 values do not vary significantly across the boroughs of Bradford, Kirklees and Wakefield, a more general "West Yorkshire English" definition of the relevant population is possible for these parameters.

In Chapter 6, productions of intervocalic /t/ and T-glottaling rates more specifically were not shown to be regionally stratified across West Yorkshire. For this reason, it is possible to provide a general description of how intervocalic /t/ is typically realised in West Yorkshire as a whole. In terms of delimiting the reference population in FSC casework, it would be sufficient to define the population at the "West Yorkshire" level rather than narrowing the population to a specific borough. However, further testing would be required to establish the magnitude by which the strength of evidence would be affected by delimiting the reference population by region or by borough. It should also be acknowledged that it is not only the regional variety that needs to be taken into account when selecting an appropriate reference population but other factors such as sex, age and social class have also been shown to affect the outcome of numerical likelihood ratios (cf. Hughes & Foulkes, 2015). The population data collected as part of this investigation relates specifically to young adult males from West Yorkshire, and although it is possible that the findings presented in this thesis may be similar in a range of other sociodemographic groups, the degree to which these results can be extrapolated across West Yorkshire is unknown.

8.2. Influence of Speech Accommodation

This section discusses the influence of speech accommodation on the phonetic parameters under investigation in this thesis as well as how social salience relates to accommodation behaviour more generally. An overview is presented of the findings in relation to the third research question addressed in this thesis and a detailed analysis of the speech of a participant who accommodated to a strong degree in both FACE and intervocalic /t/ is carried out, in order to explore within-speaker variation across a wider range of speech parameters.

RQ3: To what extent do speakers adapt their FACE and intervocalic /t/ productions across forensically-relevant scenarios involving different interlocutors and speaking styles?

A key finding of the case studies presented in Chapters 5 and 7 was that accommodation behaviour appears fairly speaker specific, with high levels of between-speaker variability present across participants. To demonstrate this, Table 8.1 presents a summary of the West Yorkshire participants' accommodation behaviour for FACE and intervocalic /t/ across Task 1 and Task 3. For this analysis, FACE accommodation behaviour is measured based on the participants' average DID value from their interlocutor during each of the paired tasks (i.e. the amount of change in the distance between interlocutor pairs based on the participants' baseline productions versus their productions in the paired task). DID values below -0.05 represented convergence and values above 0.05 represented divergence in the paired task. In cases where DID values were between -0.05 and +0.05, this was classed as maintenance.

For the assessment of accommodation behaviour in intervocalic /t/, T-glottaling rates in the PRETTY context were evaluated. As there were too few baseline /t/ productions available from Task 4, accommodation behaviour was assessed across Tasks 1 and 3 by examining differences in T-glottaling rates between the interlocutors in each pair over the course of the task. By comparing the differences in T-glottaling rates between the participant and their respective interlocutor in the first half of /t/ tokens with those of the second half of /t/ tokens, it was possible to establish whether more exposure to the interlocutor resulted in more similar T-glottaling rates. When differences in T-glottaling rates decreased by more than 5%, this represented convergence and increases of more than 5% represented divergence. Changes in differences in T-glottaling rates of less than 5% were classed as maintenance.

Table 8.1. Summary of the 30 West Yorkshire participants' accommodation behaviour in FACE and (PRETTY) intervocalic /t/ across Task 1 and Task 3.

	FACE ACCOMMODATION BEHAVIOUR						/t/ ACCOMMODATION BEHAVIOUR					
	TASK 1			TASK 3			TASK 1			TASK 3		
Participant	Convergence	Maintenance	Divergence	Convergence	Maintenance	Divergence	Convergence	Maintenance	Divergence	Convergence	Maintenance	Divergence
#006	✓			✓			✓					✓
#012			✓			✓	✓					✓
#015			✓	✓					✓	✓		
#019			✓			✓		✓				✓
#020		✓			✓				✓			✓
#021			✓			✓		✓				✓
#022			✓		✓		✓			✓		
#025	✓			✓			✓					✓
#030		✓			✓		✓			✓		
#031	✓					✓		✓		✓		
#033		✓			✓				✓	✓		
#034			✓			✓	✓			✓		
#035			✓			✓	✓			✓		
#036	✓					✓			✓			✓
#038			✓			✓	✓					✓
#040			✓		✓		✓			✓		
#041	✓			✓					✓			✓
#042		✓				✓	✓					✓
#045			✓		✓		✓					✓
#046			✓		✓				✓			✓
#048			✓			✓			✓	✓		
#049	✓			✓					✓		✓	
#050			✓			✓			✓	✓		
#054			✓		✓		✓					✓
#055			✓			✓			✓	✓		
#058			✓			✓	✓			✓		
#059			✓			✓	✓			✓		
#064	✓				✓			✓				✓
#067	✓			✓					✓		✓	
#069			✓		✓		✓			✓		
TOTAL %	26.7%	13.3%	60.0%	20.0%	33.3%	46.7%	50.0%	13.3%	36.7%	46.7%	6.7%	46.7%
N	8	4	18	6	10	14	15	4	11	14	2	14

The accommodation behaviour summary statistics presented at the bottom of Table 8.1 are also represented in the bar chart in Figure 8.2. It can be seen that the number of participants converging was higher in respect of T-glottaling rates than in FACE productions, with the highest percentage of participants converging in T-glottaling rates in Task 1. In contrast, the number of participants diverging was highest in FACE productions in Task 1, closely followed by divergence rates in Task 3 across both FACE and /t/. The number of participants who neither converged nor diverged was highest in respect of FACE productions in Task 3 and lowest in terms of T-glottaling rates in Task 3. Those participants who neither converged nor diverged and generally display less variability across different speaking situations are ideal for FSC purposes. This is because if speakers such as this are present in the known and questioned recordings, their speech is more likely to be consistent between samples meaning a stronger strength of evidence in support of the same speaker view may be reached. In this investigation, participants #020, #030 and #033 all displayed maintenance for FACE in Tasks 1 and 3 but they accommodated in terms of T-glottaling rates.

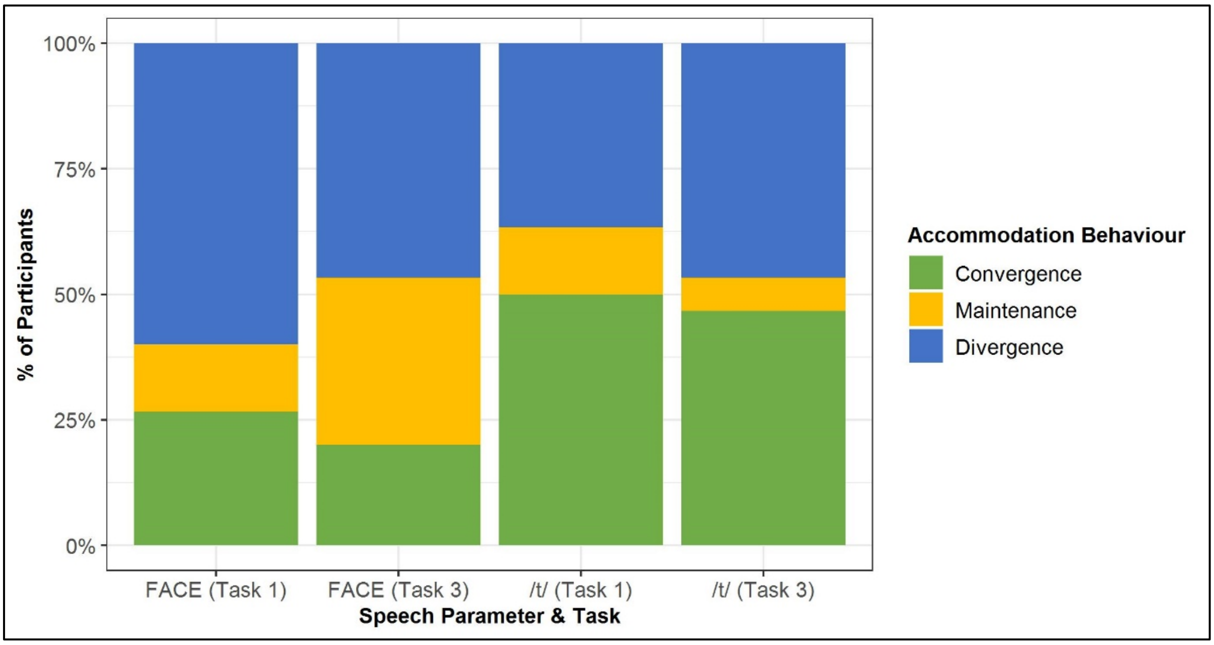


Figure 8.2. Summary of the participants' accommodation behaviour across parameters and tasks.

Overall, it was found that individual participants did not necessarily accommodate in the same way across tasks for FACE and /t/ productions. Figure 8.3 displays the proportion of participants who converged, diverged or displayed maintenance across both tasks in FACE productions and also in /t/ productions. The proportion of participants who accommodated in different ways across tasks are also represented. For FACE, most participants (63%) displayed consistent accommodation behaviour across tasks, whether that be convergence, maintenance or divergence. For /t/, the majority of participants were inconsistent across tasks (77%). It was also the case that individual participants did not necessarily accommodate in the same way across the two speech parameters under investigation. This finding is illustrated in Figure 8.4. Out of 30 participants, only 7 displayed the same type of accommodation behaviour for FACE and /t/ in each of the paired tasks (2 converged and 5 diverged consistently in Task 1, whereas 1 converged and 6 diverged consistently in Task 3).

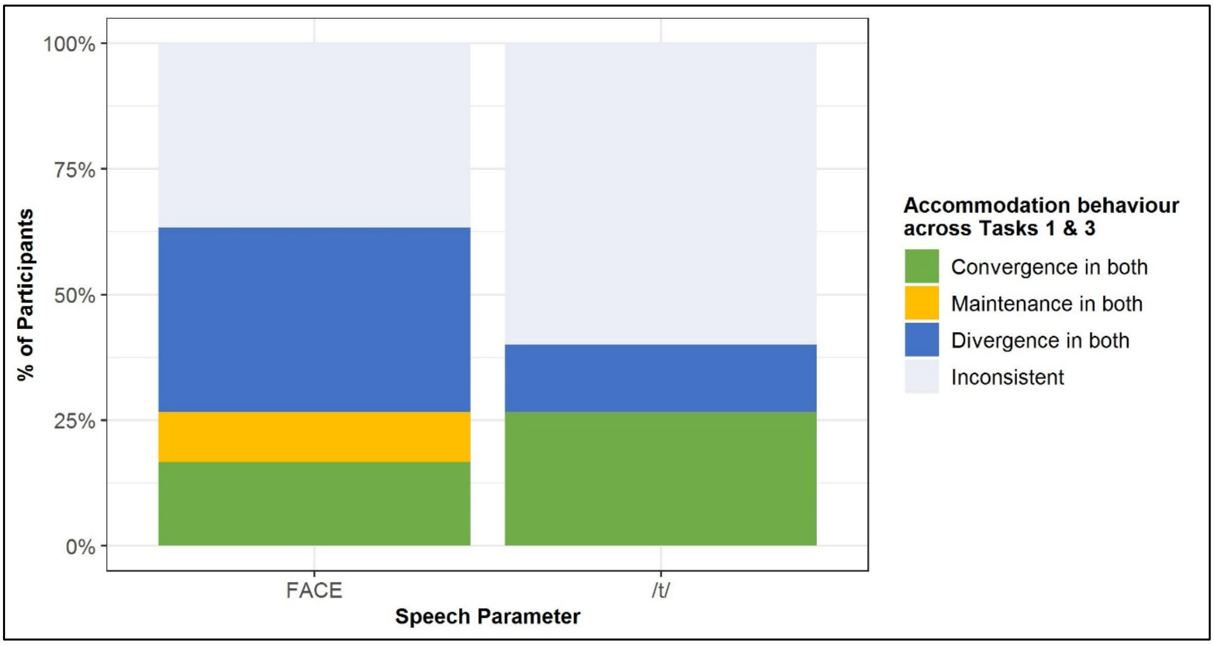


Figure 8.3. Consistency levels in the participants' accommodation behaviour across Task 1 and Task 3 for FACE and intervocalic /t/.

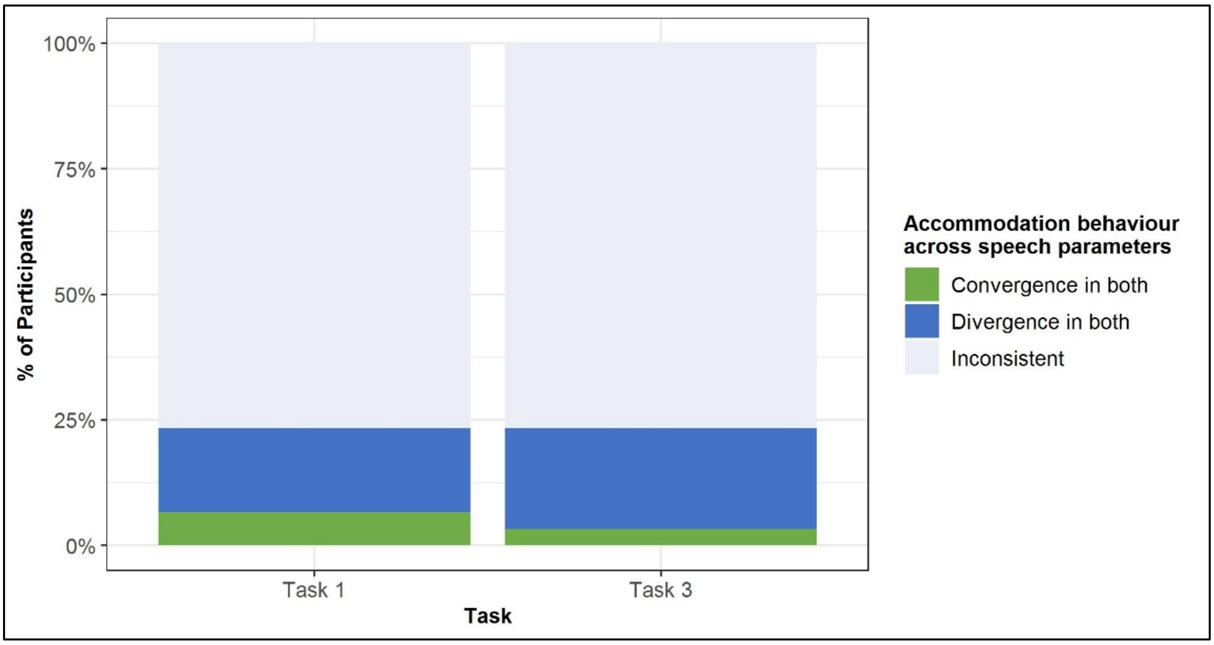


Figure 8.4. Consistency levels in the participants' accommodation behaviour across speech parameters in Task 1 and Task 3.

As seen in Table 8.1, there were no participants who displayed consistent accommodation behaviour across both speech parameters and both tasks. However, there were two sets of participants who were paired together for Task 3 who diverged in both tasks for FACE and converged in both tasks for intervocalic /t/. These pairings were participants #034 & #035 and participants #058 & #059. It is perhaps worth noting that participants #034 & #035 were a pair of friends from Wakefield and therefore may have been expected to have converged in both speech parameters during Task 3 due to being more familiar with one another. Similarly, another friend pairing from Bradford (#030 & #031) also both converged in T-glottaling rates in their paired task but in terms of FACE productions, participant #031 diverged whereas participant #030 neither converged nor diverged in this task. The fact that some of these participants diverged in respect of FACE productions is in line with the more general trend observed in this investigation whereby divergence was more common in FACE productions than in T-glottaling rates.

8.2.1. How does social salience relate to speech accommodation?

One of the primary motivations for examining the FACE vowel and intervocalic /t/ was that both of these variables were believed to be socially salient in West Yorkshire and previous speech accommodation research has shown that speakers tend to accommodate more strongly with respect to features that are salient (Cao, 2018; Smith & Holmes-Elliott, 2015; Trudgill, 1986). T-glottaling is a widely recognised feature which can elicit strong reactions based on the variable's social associations (Alderton, 2020). On the continuum of stereotypes, markers and indicators, described in Chapter 5, T-glottaling would be categorised as a stereotype of West Yorkshire speech, with this feature being commented on not only in this region but also in many other parts of the UK. Within the full WYRED dataset, a number of participants discussed this feature when describing how they spoke and one individual stated that "we don't pronounce our t's around here" when talking about the West Yorkshire region. Although this claim is not entirely accurate, it clearly indicates a strong awareness of T-glottaling being a typical feature of the West Yorkshire accent. In terms of the perceptions of this feature, participant #035 acknowledged that it is often negatively perceived by saying "when you've got all these little glottal stops and when you-, when you don't say half the words you imply, it's… it sounds stupid. But it's, it's an idiosyncratic accent and dialect and it's just the way things naturally mutate."

Monophthongal realisations of FACE in the region of [eː] and [ɛː] tend to be recognised within West Yorkshire but more generally across the UK people are perhaps less aware of this feature than they are of T-glottaling. For this reason, we may expect to find slightly less speech accommodation in FACE than in intervocalic /t/. On the continuum referenced above, the monophthongal FACE vowel would be classified as a marker of the West Yorkshire accent. An awareness of this feature is signified by the First Group's First Bus App advertisements which make use of the words *take* and *tech* being homophonous in a stereotypical West Yorkshire accent. This series of advertisements include a man with a West Yorkshire accent producing the monophthongal FACE variant [ɛ] in the slogan "Want easy travel? Tech the bus" (First Bus North, 2020). In terms of the status of this phonetic variant, there would appear to be covert prestige in the local West Yorkshire community.

During Task 3, participant #175 from Bradford described how his West Yorkshire accent had become stronger over the years: "A lot of it's now a lot more lazy. It's like as well I used to say *Facebook* [feɪsbʊk], now I just say *Facebook* [fɛːsbʊk]. It's like it's just a lot more lazy and relaxed." Although this participant does not explicitly articulate what it is that is being modified (in contrast to the "we don't pronounce our t's" comment) the direct contrast between [eɪ] and [ɛː], which can be perceived auditorily and quantified acoustically by an increase in F1, and a reduction in F2 values and vowel duration, demonstrates that this participant has some awareness of this variable being a marker of the West Yorkshire accent. In this sentence, participant #175 also produces monophthongal FACE tokens in the words *lazy* [ɬe̞ːzɪ] and *say* [sɛː].

Overall, it would appear that the participants analysed in this investigation accommodated in both FACE and /t/ productions. There tended to be higher levels of convergence in T-glottaling rates than in FACE productions across both paired tasks, whereas there were higher levels of divergence and maintenance in FACE productions. Although West Yorkshire speakers appear to be aware that the short monophthongal FACE variant [ɛ] is strongly associated with the area, the levels of speech accommodation in FACE productions across tasks tended to be more subtle and less noticeable than the variability in /t/ productions. One reason for this could be that T-glottaling is a stereotype rather than a marker and usually tends to elicit consistent style-shifting effects. Furthermore, while FACE realisations varied across the three West Yorkshire boroughs, T-glottaling rates did not.

Figure 8.5 illustrates the amount by which each participants' average FACE F2 midpoint values and T-glottaling rates changed from Task 1 to Task 3. This visualisation is intended to demonstrate: (1) the range of between-speaker variation in terms of within-speaker variability across tasks, and (2) how the two parameters relate to one another. Task 4 data is not represented in this figure as there was limited data available to be able to reliably evaluate /t/ productions in this task. The percentage change in FACE F2 values, as opposed to F1 or F3, is plotted because this was the only formant that was found to vary significantly across tasks in Chapters 4 and 5. Average F2 values in raw hertz were used to calculate percentages rather than BDM transformed values in order to present the data in a way that is more easily interpretable.

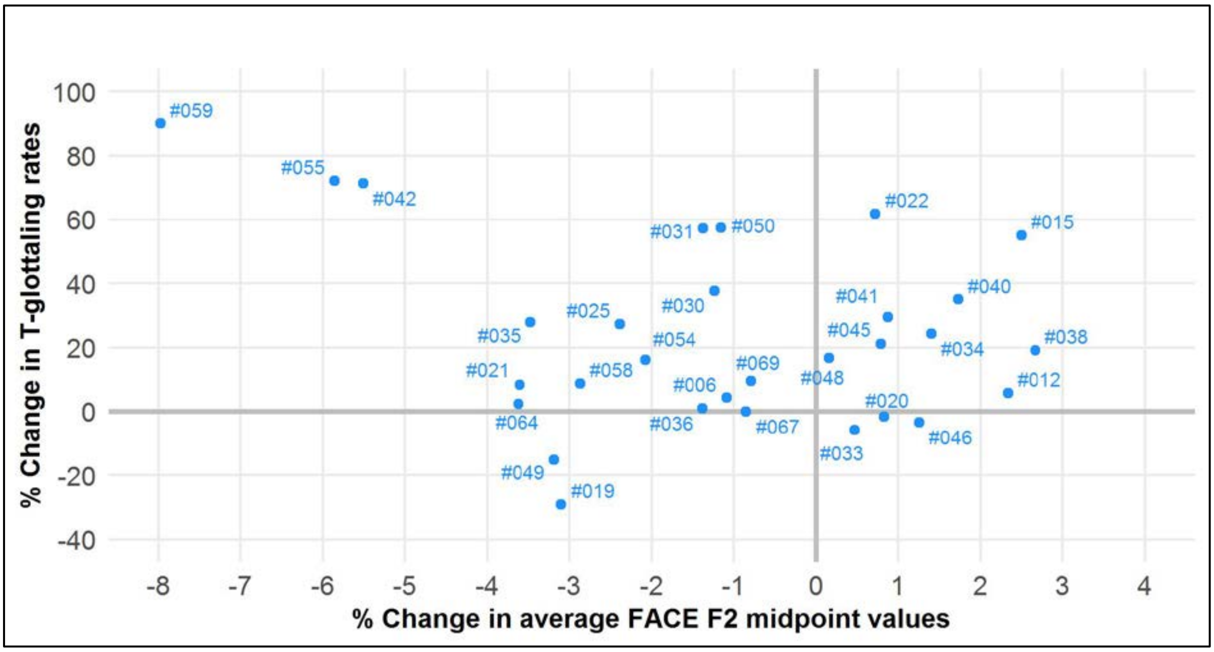


Figure 8.5. Percentage change from Task 1 to Task 3 in FACE F2 values and T-glottaling rates.

Overall, the results shown in Figure 8.5 indicate that there was a general tendency for FACE F2 values to be higher in Task 1 than in Task 3, and for T-glottaling rates to be higher in Task 3 than in Task 1. In order to determine whether or not the participants' change in average FACE F2 midpoint values were correlated with their change in T-glottaling rates from Task 1 to Task 3, a Spearman's rank correlation test was conducted. The results of this test showed that there was no correlation between these values ($r_s = -0.103$, $n = 30$, $p = 0.588$). This indicates that it is not possible to reliably predict accommodation behaviour in one of these two speech parameters based on that of the other. However, with reference to Figure 8.5, it can be seen that many of the participants displaying the smallest change in FACE F2 values from Task 1 to Task 3 also had relatively consistent T-glottaling rates across tasks. Furthermore, if we focus on the results of those participants whose FACE F2 values decreased from Task 1 to Task 3, it can be seen that those with the most extreme decrease in F2 from Task 1 to Task 3 also had some of the largest increases in T-glottaling rates across these tasks. It is those participants on the outskirts of this plot that have the potential to be most problematic in the context of FSC casework, as these high levels of within-speaker variability could possibly contribute to evidence in support of a different speaker view, despite the separate samples being produced by the same speaker. For example, participants #059, #055, and #042 could all possibly pose a problem for a forensic expert, especially if high levels of within-speaker variation are also present in other speech parameters.

8.2.2. Case study of participant #059

As stated in the previous section, participant #059 was one of the participants who accommodated the most in terms of both FACE and intervocalic /t/ productions. In Table 8.1, it was shown that this participant diverged in respect of FACE productions in Task 1 and Task 3 whereas he converged in terms of T-glottaling rates in both of these tasks. In Figure 8.5, participant #059 can also be seen to display the highest amount of change from Task 1 to Task 3 in both speech parameters. His T-glottaling rate increased from 10% in Task 1 to 100% in Task 3 and his average FACE F2 value decreased by 8% (142 Hz) from Task 1 to Task 3. This sub-section presents a case study of participant #059's variability across speaking tasks involving different speaking styles and interlocutors, in the form of a fictional FSC case. The purpose of this is to examine the extent to which a range of other speech parameters varied. It is possible that high levels of variation across tasks may be present in parameters other than FACE and intervocalic /t/. If this is the case, it will be useful to explore whether or not those features that are most socially salient are the ones that vary the most.

In line with methods commonly applied in FSC analyses in the UK (Gold & French, 2011, 2019), a combined auditory and acoustic approach was taken to analyse both segmental (vowel and consonant realisations) and supra-segmental (pitch and voice quality) parameters in the speech of participant #059 recorded during three tasks. These specific parameters were chosen as they are all routinely analysed in FSC casework. Additional features such as intonation and tempo were not analysed in this case study as participant #059 was not deemed to display unusual intonation patterns or have a perceptually very high or low articulation rate across the paired tasks. As articulation rate in particular has been found to have low discriminatory power except for in instances of very high or low rates (Gold, 2014, p. 33), this feature of participant #059's speech would be unlikely to be analysed in a real FSC case. Furthermore, as the answer message task required the participants to convey as much information as possible within three minutes, a higher articulation rate was expected in this task.

Firstly, an assessment of long-term f_0 was carried out. To conduct this analysis, f_0 edit files containing only the isolated speech of participant #059 were produced manually using Sony Sound Forge (version 9.0e). These edit files were subsequently analysed using a Praat script (Harrison, 2019a) designed to automatically calculate and display the f_0 of a selected audio file and calculate the f_0 distribution. Within the Praat script, suitable settings were specified in order to obtain reliable f_0 measurements (please see Appendix 7 for details). Across all three tasks, the participant spoke at a normal speaking level and his f_0 distributions were slightly positively skewed in line with this. Table 8.2 provides summary statistics in relation to the participant's f_0 across the three tasks. Overall, participant #059's f_0 results were very consistent across tasks and did not appear to vary as a result of differing speaking styles or the influence of his interlocutors across the two paired tasks.

Table 8.2. Participant #059's f_0 summary statistics across tasks.

	f_0 Statistics (Hz)			
Task	**Average**	**SD**	**Median**	**Alterative Baseline**
1	113	12	111	101
3	114	11	112	104
4	110	9	110	101

Next, voice quality was analysed auditorily using a modified version of the Vocal Profile Analysis (VPA) scheme (Laver, 1980) which is currently employed at J P French Associates and Soundscape Voice Evidence, which are the two main forensic speech laboratories in the UK. For a detailed description of the modified VPA scheme used for this analysis and the various voice quality features that are captured using this scheme, please see Wormald (2016, sec. 4.4). Participant #059's voice quality was considered to be very similar across the three different tasks, both at a componential and holistic level. His voice quality was characterised by the following combination of features:

- Advanced tongue tip/blade and slight sibilance
- Raised and tense larynx
- Creaky voicing with slight, intermittent harshness

Although levels of creakiness varied across tasks, with more creakiness during the casual paired conversation than in the other tasks, overall the participant's voice quality did not appear to vary substantially depending on speaking style or interlocutor.

In order to examine participant #059's vowel realisations acoustically, midpoint F1~F3 values from multiple tokens of a range of mostly monophthongal vowel phonemes were measured and logged using a Praat script (Harrison, 2019b). The vowel phonemes analysed are discussed with reference to Wells' (1982) lexical sets. For this specific analysis, vowel sets such as FACE and GOAT, which would typically be realised as diphthongs in Standard English, were only measured if the specific vowel token was perceived auditorily as a monophthong and the formants were relatively stable across the vowel. Figure 8.6 contains vowel plots for participant #059 across each of the three tasks. In this figure, the dots represent the midpoint F1 and F2 values of each individual vowel token measured, whereas the labelled boxes indicate the average midpoint F1 and F2 value for each vowel set. The dots are colour-coded in line with the labels for each vowel set so that tokens of each vowel phoneme can be distinguished from one another.

When comparing the vowel space scatterplots from each task, it can be seen that Task 1 and Task 4 look broadly similar whereas Task 3 appears to be comparatively more compressed. The vowel space observed in Task 3 appears relatively typical of a West Yorkshire speaker and may indicate that participant #059 produced more local vowel realisations during this task. There are also a few slight differences in the positions of some of the average F1 and F2 labels for certain vowel sets across tasks, however, it should be noted that in some cases there were only a few tokens of each vowel set available. Furthermore, as noted in Chapter 4, the phonetic environment can play an important role in vowel realisations and therefore differences in the phonetic environment of tokens between tasks may have affected the average values. For example, realisations of the JUICE vowel appear to vary quite drastically across tasks, however, there were only two tokens available in Task 1 (*Eugene* and *U*) and only one token in each of Tasks 3 and 4 (*music* and *Eugene*, respectively).

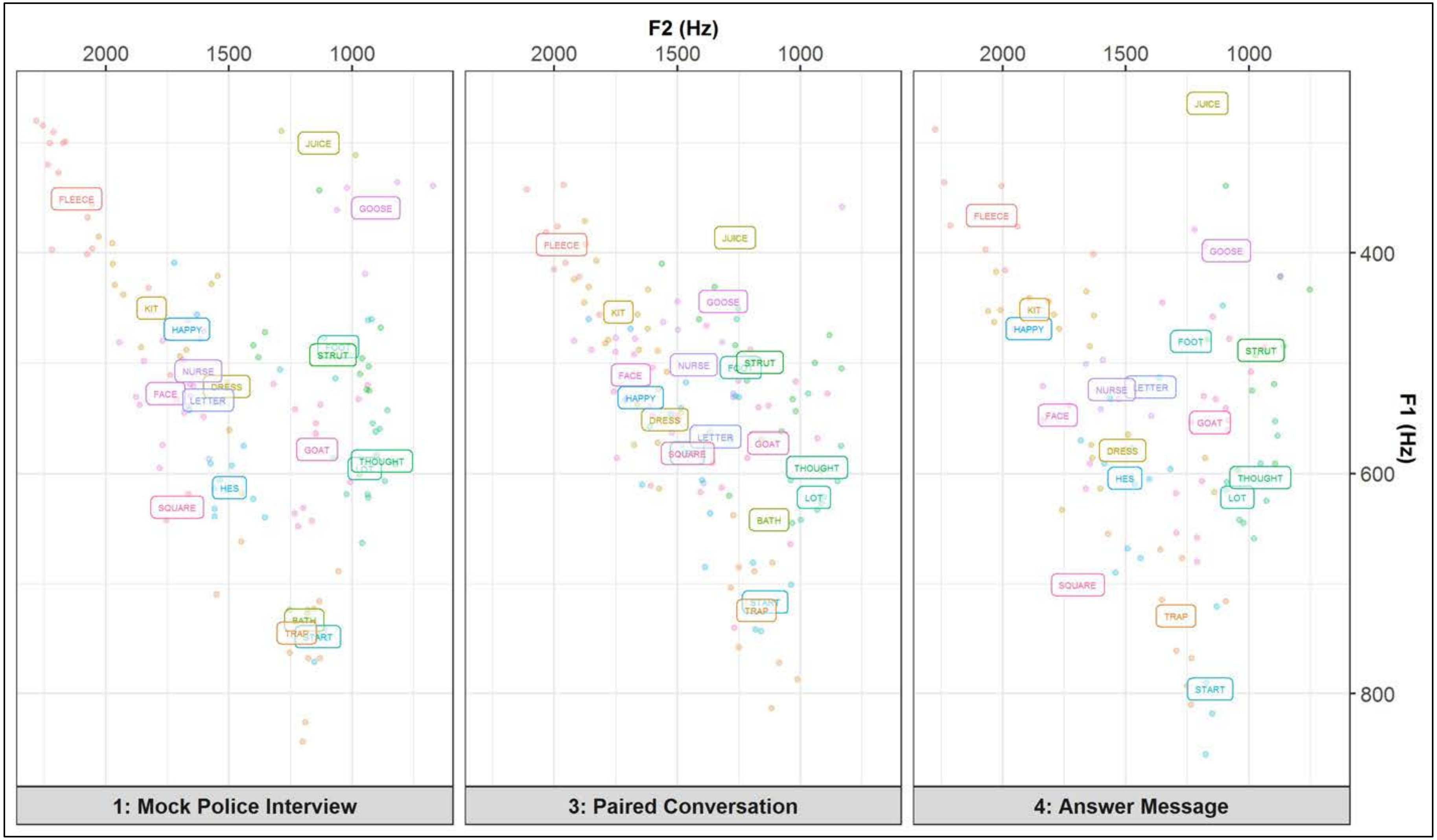


Figure 8.6. Vowel space of participant #059 across tasks.

The midpoint F3 distributions from all of the vowel tokens measured from each of the tasks are presented in the form of a boxplot in Figure 8.7. Summary statistics relating to the F3 values are also provided in Table 8.3. These findings reflect that, overall, F3 values do not shift to a large extent across the tasks.

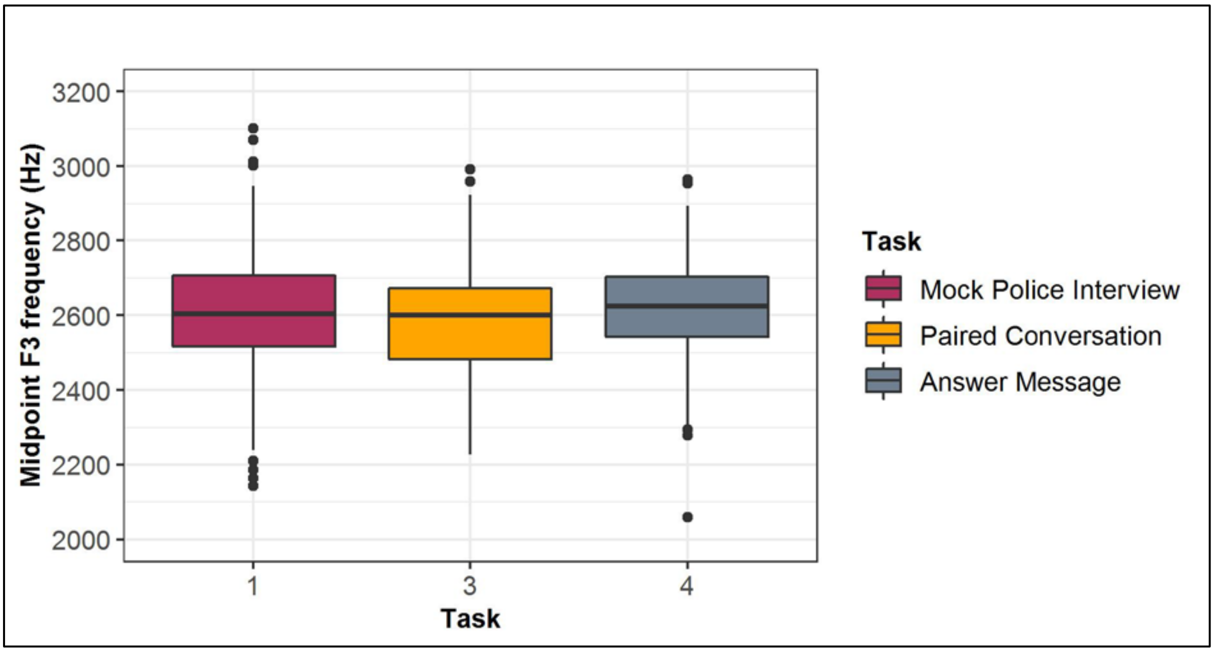


Figure 8.7. Participant #059's F3 distributions across tasks.

Table 8.3. Participant #059's F3 summary statistics across tasks.

	F3 Statistics (Hz)		
Task	**Average**	**Median**	**SD**
1 (N=129)	2613	2610	179
3 (N=117)	2576	2577	155
4 (N=121)	2608	2623	143

The final part of this case study involved analysing participant #059's vowel and consonant realisations auditorily. Table 8.4 provides an overview of the range of variables that were analysed and a corresponding IPA transcription of the main realisations observed across each task. Where multiple variants are transcribed, there was some variation across different tokens within an individual task. In these cases, the predominant variant is presented first.

Table 8.4. Summary of participant #059's vowel and consonant realisations across tasks.

Variable	TASK 1	TASK 3	TASK 4
FLEECE	[iː]	[iː]	[iː]
KIT	[ɪ]	[ɪ]	[ɪ]
DRESS	[ɛ]	[ɛ]	[ɛ]
TRAP	[a]	[a]	[a]
BATH	[a]	[a]	*No tokens*
START	[ɑː]	[ɑː]	[ɑː]
STRUT	[ʊ]	[ʊ̞]	[ʊ]
THOUGHT/NORTH	[ɔː]	[ɔː]	[ɔː]
LOT	[ɒ]	[ɒ]	[ɒ]
FOOT	[ʊ]	[ʊ]	[ʊ]
GOOSE	[ʉː]	[ʉː]	[ʉː]
JUICE	[jʉː]	[jʉː]	[jʉː]
Hesitation: *er & erm*	[ɜː ~ ʌː]	[ɜː ~ ʌ]	[ɜː ~ ɜ̠ː]
happY	[ɪ]	[ɪ]	[i̞]
LETTER/COMMA	[ə]	[ə]	[ə]
NURSE	[ɜː]	[ɜː]	[ɜː]
PRICE	[aɪ]	[aɪ ~ aː]	[aɪ ~ a]
FACE	[eː ~ ɛː ~ eɪ]	[eː ~ ɛː ~ eɪ ~ ɪ]	[eː]
GOAT	[ɔː]	[ɔː]	[ɔː]
MOUTH	[aʊ]	[aʊ ~ a]	[aʊ ~ ɒ]
Word-medial /t/	[t]	[ʔ ~ t ~ s]	[ʔ]
Word-final /t/	[ʔ]	[ʔ ~ t^s]	[ʔ ~ t ~ Ø]
Word-initial /ð/	[ð]	[ð]	[ð ~ d ~ z]
Word-medial /ð/	[v]	[v]	[v]
/θ/ (all positions)	[θ ~ f]	[f ~ θ ~ s]	[θ ~ f]
Pre-vocalic /ɹ/	[ɹ ~ ʋ]	[ɹ]	[ɹ ~ $ɹ^ʋ$]
Word-initial /l/	[ɫ]	[ɫ]	[ɫ]
Syllable-initial /h/	[h ~ Ø]	[h ~ Ø]	[h ~ Ø]
/k/ (word-initial and final)	[k^h ~ k]	[k^h ~ k]	[k^h ~ k]
ING-**suffix**	[n]	[n]	[n]

It can be seen that the majority of features were produced consistently across tasks. However, there were some differences in respect of certain vowel and consonant realisations. The distribution of variants for FACE, PRICE, MOUTH and the hesitation markers *er* and *erm*, all varied to some degree across tasks but nevertheless the predominant variant tended to be consistent overall. There was also subtle variation in the realisation of /ɹ/ and word-initial /ð/ across tasks but again the predominant variant was present across all three tasks.

The most notable type of variation across tasks was in the proportional use of T-glottaling, TH-fronting and H-dropping. Interestingly, all of these variables are generally considered to be socially salient. Although an in-depth analysis of T-glottaling rates accounting for phonetic context was beyond the scope of this case study, general trends were observed in word-medial and word-final /t/ across tasks. As previously described above, for word-medial /t/, use of [t] was much more common in Task 1 than in Tasks 3 and 4, where [ʔ] was the dominant form. For word-final /t/, [ʔ] was the most prevalent realisation across all three tasks, however, examples of [t^s] and [t] were also found in Tasks 3 and 4 respectively.

TH-fronting was observed across all three tasks in /θ/ and word-medial /ð/. For word-medial /ð/, participant #059 categorically used [v] in all tasks. Whereas for /θ/, realisations varied by word position and by task. Figure 8.8 presents the percentage of /θ/ tokens that were produced as [f] across tasks in word-initial, word-medial and word-final contexts. It can be seen that TH-fronting rates vary across tasks to quite a strong degree, with use of [f] being highest in Task 3 and lowest in Task 1. This is in line with the style-shifting trends we might expect, whereby the non-standard variant is more prevalent in the more casual setting. It was also found that participant #059 pronounced /θ/ as [f] more often word-medially and word-finally than word-initially. However, these findings are based on very few tokens and must therefore be treated with caution. It should also be noted that no tokens of word-medial or word-final /θ/ were available for analysis in Task 4.

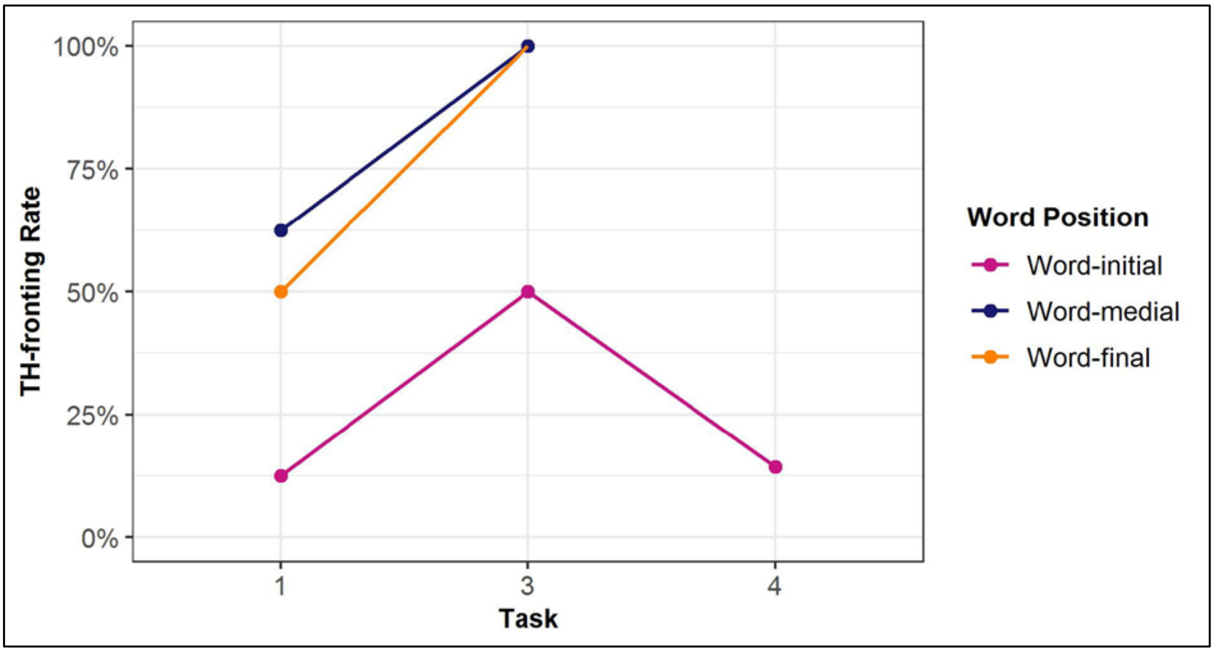


Figure 8.8. Participant #059's voiceless TH-fronting rates across contexts and tasks.

Participant #059 also displayed varying levels of H-dropping across the three tasks. Figure 8.9 shows the percentage of tokens of syllable-initial /h/ which were realised as Ø, separated by word type across tasks. It can be seen that in all three tasks, H-dropping rates were much higher in function words than in content words. In Task 1, 28 of the 79 /h/ tokens analysed were from keywords that the participant had to produce as part of the mock police interview task. Although these words were spoken as part of a free-flowing conversation, they were shown on screen to assist the participant in answering the police interviewer's questions, and largely included proper nouns (such as *Hatfield* and *Hogan*), therefore they were considered likely to be more closely aligned with read speech than the other content words produced during this task. For this reason, keywords were separated from the other content words and, as can be seen in Figure 8.9, participant #059's H-dropping rates varied drastically across these two categories.

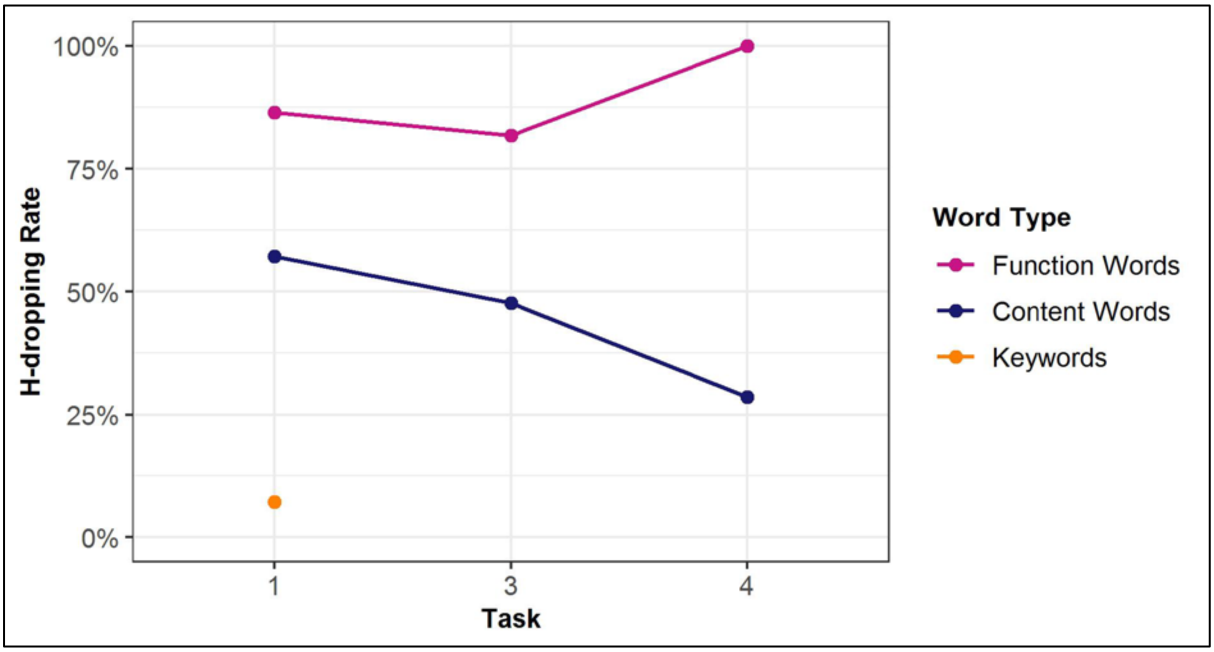


Figure 8.9. Participant #059's H-dropping rates across contexts and tasks.

Overall, H-dropping rates in both content and function words were slightly higher in Task 1 than in Task 3. Although the opposite pattern was expected due to H-dropping being a somewhat stigmatised feature, these differences were relatively small. Furthermore, when including the keywords from Task 1 within the "Content Words" category, the H-dropping rate decreased from 57% down to 24%, well below the H-dropping rate observed in content words in Task 3. In Task 4, H-dropping rates were the highest of the three tasks in function words and the lowest of the three tasks in content words, however, there were only 10 tokens of syllable-initial /h/ available in this task and therefore the results should be treated with caution.

Across participant #059's three speaking tasks, there were strong similarities in terms of f_0, voice quality, and a range of vowel and consonant realisations. However, there were some differences across tasks in the distribution of socially salient features such as T-glottaling, TH-fronting, H-dropping, as well as auditory and acoustic differences in FACE productions. On balance, if the three speech samples examined in this analysis were to be compared in a real FSC case (in which the identity of the speaker was unknown), the levels of variability across samples would be unlikely to be interpreted by the expert as providing evidence in support of the different speaker view. However, some of the differences may have an impact on the strength of the evidence in support of the same speaker view.

While the findings of this case study are perhaps reassuring from a FSC perspective, it must be recognised that the recordings used for this analysis were all of studio-quality and each contained plenty of free-flowing speech which meant that a wide range of speech parameters could be assessed. However, in the vast majority of FSC cases the samples being compared will be of much poorer quality than those analysed in this investigation. Samples are often mismatched in terms of channel and technical quality, and in many instances there is limited free-flowing speech available for analysis. As a consequence of this, it is sometimes not possible to obtain reliable vowel formant measurements or to assess other elements such as voice quality. If it had been the case that only a subset of participant #059's vowel and consonant realisations could reliably be analysed auditorily, the variability across samples in the socially salient features would have had a greater impact on the results and could have resulted in a weaker positive conclusion or potentially even a negative conclusion. Consequently, it is in those cases where only limited forms of analysis can be undertaken where the influence of speech accommodation and differences in speaking style may be of most concern.

8.3. Implications and recommendations for forensic speech science

The first part of this section addresses the fourth research question posed in this thesis. In the second part of this section, recommendations and considerations for FSC caseworkers and researchers in the field of LADO are outlined.

RQ4: What are the potential implications of speech accommodation for FSC casework?

The results of the case studies presented in Chapters 5 and 7 have shown that the West Yorkshire participants accommodated in FACE and intervocalic /t/ productions across speaking tasks. As both of these speech parameters are routinely examined as part of a FSC analysis, the findings of this study have direct implications for FSC casework involving West Yorkshire speech. Although the magnitude of accommodation varied greatly between participants, there were some participants who accommodated to such a degree that their levels of variation might have had an impact on the outcome of a FSC case, had their identity been unknown. These findings indicate that speech accommodation has the potential to cause certain speech parameters to vary within an individual to such an extent that an expert might misinterpret differences between samples as providing evidence in support of the different speaker view when they can in fact be accounted for by speech accommodation.

The findings of this investigation have highlighted that it is not easy to predict accommodation behaviour as it appears to be highly speaker specific. In this investigation, examples of convergence, divergence and maintenance were observed across participants for both parameters during the mock police interview scenario and the casual conversations. The group results suggest that accommodation behaviour is not necessarily consistent across tasks and even less consistent across speech parameters. However, the results of this investigation suggest that in forensically-relevant speaking situations, accommodation is most likely to occur within socially salient variables.

As shown in Figure 8.2, there were less instances of maintenance in T-glottaling rates (the more salient of the two variables) than in FACE productions. Results of the case study presented in Section 8.2.2 further indicate that it is the more socially salient variables for which higher levels of within-speaker variability are present. This suggests that experts should exercise particular caution when examining socially salient variables compared to the ones that are below the level of consciousness and be aware that the salient variables may be less consistent within an individual. It is therefore crucial that experts have a high level of awareness of a wide range of accents so that they can accurately identify those features that are socially salient in any given accent. Although FACE is salient in West Yorkshire English, this is not the case in many other varieties of English and consequently accommodation effects for this variable are likely to be stronger in speakers from West Yorkshire than speakers from other regions.

The results of this investigation also indicate that more caution may be required when there are more extreme mismatches between samples in terms of speaking styles and when the respective interlocutors across samples are very distinct. For example, in the case of FACE, participants were found to accommodate to a stronger degree in the mock police interview than in the casual paired conversation. One reason for this could be that the interview scenario is more likely to evoke a phonetic response resulting in deviations from a speaker's usual way of speaking, perhaps resulting in more standard forms. Another explanation could be that participants were more relaxed in the paired conversation and more closely matched to their respective interlocutors, in terms of accent and gender than they were in the police interviews. The practical implication of this finding for FSC casework is that more extreme mismatches across samples (in terms of interlocutors and speaking styles) may result in higher levels of accommodation overall. In practice, if a known speaker were interviewed by a police officer with a different accent to their own this could potentially lead to changes in their speech that would not be observed if they were interviewed by an officer with a similar accent.

8.3.1. Recommendations

8.3.1.1 FSC casework

Overall, the findings of this study provide evidence to indicate that speech accommodation deserves just as much recognition as other potentially problematic factors, such as mismatches in speaking mode and channel, and the effects of non-contemporaneity, drug use or alcohol intoxication on the speech in the evidential samples. The results presented in this thesis highlight where caution needs to be taken, in terms of the types of speech parameters that are most heavily influenced by accommodation as well as the speaking situations where accommodation may be more likely to occur. The high levels of between-speaker variability in accommodation behaviour also reinforce the need for experts to be aware of the possible consequences that speech accommodation may have on a person's speech. Consequently, it is recommended that any potential effects of speech accommodation should be taken into account as part of a caseworker's standard operating procedures.

8.3.1.1.1 Reference samples

When sourcing reference samples for FSC casework the instructing party should aim to gather as much reference speech as possible, ideally from a range of different speaking styles, so that the expert can adequately consider the within-speaker variation of the known speaker. However, in practice, police interviews account for around two thirds of known samples in UK FSC cases, and in many cases this is the only source of reference material available. Consequently, wherever possible, experts should assign more weight to (sections of) known samples that most closely match the questioned sample in terms of speaking style, vocal effort, technical quality and channel. In cases where extreme mismatches exist between the known and questioned samples, the expert must determine whether or not the comparison can be undertaken. In terms of speech accommodation, greater mismatches across samples in terms of formality, familiarity and accents of the respective interlocutors may result in higher levels of within-speaker variability.

8.3.1.1.2 Assessing suitability

Although it is most likely not necessary to routinely analyse the speech of any interlocutors present in forensic samples, it may be useful to take their speech into account and to consider the surrounding context of the interaction when completing case-by-case assessments of suitability. In some cases, this may risk exposing the expert to biasing contextual information and therefore this assessment could be carried out by another trained analyst with any relevant information being passed to the expert, as and when required. In the same way that factors such as alcohol intoxication, stress, and drug use have been shown to influence a range of speech parameters in inconsistent ways across speakers (cf. Braun & Künzel, 2003; Kirchhübel, Howard, & Stedmon, 2011; Papp, 2009, respectively), the consequences of speech accommodation and audience design are not straightforward. Nevertheless, if speaking styles are highly dissimilar across the evidential samples or there is evidence to suggest a speaker is strongly adapting their speech as a result of who they are speaking to, then it may be necessary to deem the samples unsuitable for comparison.

8.3.1.1.3 Analysis and evaluation of evidence

In FSC cases where the analysis results differ between the known and questioned samples, it could be worthwhile for an expert to review their findings after carrying out a brief assessment of the speech of any interlocutors present in the recordings, if they have not done so already. This would enable them to consider whether any differences could possibly be explained by accommodation effects. When considering any potential effects of speech accommodation during a FSC analysis, it is recommended that the expert pay particular attention to socially salient variables and consider how consistent they are both within and between samples. In order to be able to identify socially salient variables, it is important that the expert profiles the accent of the speaker as fully as possible. In instances where poorer quality samples are compared, and the range of features that can be analysed are reduced, the consequences of accommodation may be more extreme and therefore a more conservative approach should be taken. In cases where high levels of variability are observed across known and questioned samples, this is likely to result in a weaker positive conclusion or a negative decision (unless stable and distinctive features are also present across the samples).

8.3.1.2 LADO

The results of the present investigation may also have implications for disputed nationality cases as well as for LADO research more broadly. As discussed in Section 2.2.2.2, there is some debate as to who should carry out language assessments for applicants involved in disputed nationality cases. Based on the findings of this study, it seems plausible that an applicant might accommodate during a language assessment as a consequence of the input they receive from their interlocutor and therefore this might provide motivation for assessments to be conducted by native speakers of the language the applicant claims to speak. Although this investigation did not assess the participants' ability to imitate their interlocutor, there were many examples of the participants' FACE and /t/ realisations becoming more similar to those of their interlocutor during the paired tasks.

In this study, participants were largely found to accommodate more strongly during the mock police interview than in the casual paired conversation, with higher levels of divergence being observed than convergence. Based on this, we can hypothesise that an applicant may be more likely to accommodate during a language assessment conducted by an analyst who does not share the same language background as themselves compared to an analyst who does have the same language background. Furthermore, the consequences of accommodation may be more extreme when language backgrounds differ as opposed to accents and dialects of the same language. For this reason, it may be advisable for applicants to be interviewed by a native speaker of the language that the applicant claims to speak, in order reduce the amount that they might deviate from their typical way of speaking. However, regardless of who conducts the assessment, it is important to recognise that the applicant may converge towards or diverge away from their interlocutor and this should be considered when determining the outcome of the assessment. It is therefore recommended that in any assessments of this kind, it may be beneficial to first consider whether there is any indication that the speaker is disguising their speech, and if there is no evidence of this, potential effects of speech accommodation should be taken into account.

8.4. Limitations

This section acknowledges some of the limitations associated with the research presented in this thesis. These include issues related to the nature of the semi-spontaneous speech data analysed in this investigation, as well as limitations related to the methods of analysis that were employed.

One aim of this investigation was to examine the extent to which speakers from West Yorkshire accommodate in their productions of the FACE vowel and intervocalic /t/, whilst also considering within-speaker variability more generally in the context of forensically-relevant scenarios. In order to be able to do this effectively, it was necessary to deal with spontaneous speech elicited during conversational tasks with forensic relevance. The WYRED database was designed with forensic applications in mind and was deemed to be suitable for use in this investigation as it included speakers from a homogeneous population completing a range of varied speaking tasks. However, one of the main limitations of using conversational data which had not been elicited for the specific purpose of investigating phonetic accommodation was that the influence of interlocutor and style were conflated in this study and could not be separated from one another. For this reason, the term “accommodation” was defined more broadly in this study to refer to changes that occur across multiple speaking situations involving different interlocutors. While this made the investigation into speech accommodation more challenging, it also made the findings of this study more forensically relevant, as authentic samples examined in FSC casework often involve mismatches in terms of speaking style, interlocutor and channel.

Another implication of analysing speech elicited in a conversational setting was that the input from the interlocutor could not be controlled in the same way that a model talker's speech can usually be controlled and manipulated in non-interactive speech-shadowing experiments. In each of the paired tasks, both the participant and their respective interlocutor had the opportunity to accommodate over the course of the task. However, it is likely that this was less of an issue for the mock police interview task than in the casual paired conversations, as the researcher in Task 1 played the role of a police interviewer and largely asked the same kinds of questions in every interview that took place. The interlocutors in Task 3, on the other hand, had a spontaneous casual conversation. Furthermore, the interlocutor pairs in the Task 3 conversations were both participants and they were aware that they had to hold a conversation with their partner for twenty minutes, and therefore they may have been more likely to accommodate in order to maintain a positive interaction. Whilst it is acknowledged that this meant the study was less controlled than previous speech accommodation studies involving non-interactive data, it also meant that the data being analysed was more natural and consequently the findings of this study are considered to be more applicable in real-world situations.

With regards to the FACE accommodation analysis, the FACE productions of each participant's respective interlocutors in Tasks 1 and 3 had to be represented using one average measure of their F1 and F2 values. A disadvantage of this method of representing the interlocutors' speech was that this meant FACE was effectively treated as being constant across the task and the interlocutors' full range of FACE realisations was not captured. It is likely that the interlocutors varied their productions across each conversation and may have in fact converged towards, or diverged away from, the participant. An alternative way to represent the interlocutors' realisations, would have been to calculate the distance of each participant's first FACE token from their interlocutor's first FACE token and then the second and so on. This method would work well in a speech-shadowing task where each time the model talker produces a specific token the participant would repeat that same token in quick succession. However, due to the spontaneous nature of the data in this study, FACE tokens were only available within each interlocutor's speech when they happened to occur naturally in the conversation. It was not the case that matching FACE tokens were produced in parallel between the pairs of speakers in these tasks. Consequently, it would have been slightly arbitrary to have taken this approach because this would have resulted in tokens from different phonetic environments being compared with each other, with varying time lapses between the pairs of tokens being produced. Instead, it was considered more logical to compare each participant's individual FACE tokens with one consistent holistic measure of their interlocutor's speech. Similar methods to this have also been used in previous phonetic accommodation studies such as Cao (2018).

Due to the spontaneous nature of the two paired speaking tasks, it was not possible to control exactly when the participants and their respective interlocutors produced FACE and intervocalic /t/ tokens, nor was it possible to specify the words in which these tokens were produced (although, there were a number of keywords that typically occurred across Tasks 1 and 4, as outlined in Chapter 3). One potential limitation of this was that the phonetic environments in which FACE and intervocalic /t/ tokens occurred were not consistent between participants or within participants across tasks. This meant that some differences in productions may have resulted in part from the phonetic environment, rather than speaking task or accommodation effects. Although it was possible to account for this in the statistical models when assessing regional variability, this could not be factored into the analysis of FACE DID values when examining accommodation behaviour.

A further limitation was that in respect of word-medial, intervocalic /t/ specifically, there were relatively few tokens available for analysis. This was especially true in the case of the Task 4 data which had to be excluded from the main analysis due to the scarcity of /t/ tokens. Although the total number of tokens available per participant was considered sufficient for the investigation of regional variation, ideally more tokens would have been analysed from each individual task when examining accommodation behaviour. However, the semi-spontaneous nature of the data analysed in this thesis was considered to be reflective of the type of data that is typically available when analysing naturally occurring speech for purposes including FSC casework. Although this study had slightly lower levels of experimental control than many previous speech accommodation studies, it was also considered to have higher levels of ecological validity.

The final point to raise in this section is that both of the paired tasks involved face-to-face conversational interactions. Although this is not necessarily a limitation, previous research has suggested that speakers may be more likely to adapt their speech to aid communication in situations where they are unable to see one another (Fitzpatrick et al., 2015; Hazan et al., 2019). One way in which this could have been explored in this study would have been by analysing the participants' WYRED Task 2 data, where they each had a telephone conversation with a researcher who was not inside the sound booth and could not be seen. However, as explained in Chapter 3, a decision was taken not to analyse the Task 2 recordings as I was the researcher who took part in these conversations. It is not known if or how the participants in this study made use of any non-verbal strategies during the paired tasks, such as gesture and gaze, and it is not possible to say whether their accommodation behaviour would have been more extreme had they been reliant solely on speech to communicate with their interlocutor. However, the overall aim of this investigation was to examine within-speaker variability in FACE and intervocalic /t/ in the context of forensically-relevant scenarios, and by analysing the face-to-face WYRED data, this goal was achieved. Despite the limitations highlighted above, it is considered that the findings of this study can be used to inform the fields of forensic speech science and sociophonetics.

8.5. Further Work

The research presented in this thesis has provided inspiration for many further avenues of study. Some ideas for further research are detailed below.

8.5.1. Examine more speech parameters

This investigation of regional variation and accommodation has focussed on two specific features of West Yorkshire English speech: the FACE vowel and intervocalic /t/. In line with previous accommodation studies which have found that speakers tend to accommodate more strongly with respect to socially salient features (Cao, 2018; Smith & Holmes-Elliott, 2015; Trudgill, 1986), many participants in the present study accommodated in FACE and intervocalic /t/ productions. Participant #059 displayed particularly high levels of convergence in T-glottaling rates and divergence in FACE productions. Furthermore, the case study presented in Section 8.2.2 revealed that participant #059 also had high levels of variability between tasks in relation to two further socially salient variables: H-dropping and TH-fronting. Based on these findings, an idea for future work would be to explore accommodation in a range of other features that are considered to be socially salient within this specific speech community.

Another idea for future research would be to examine regional variability by using other acoustic modelling techniques such as by measuring Mel-frequency cepstral coefficients (MFCCs). In contrast to phonetic features, MFCCs are abstract properties of the acoustic signal which reflect various aspects of the vocal tract in a holistic way. Brown & Wormald (2017) describe them as "short-term spectral features that take the log of the magnitude spectrum, which is then Mel-filtered to approximate the shape of the vocal tract at the time the signal is produced" (Brown & Wormald, 2017, p. 426). Spectral features such as this are frequently used in automatic speaker recognition systems (e.g. VOCALISE; Alexander, Forth, Atreya, & Kelly, 2016; and Y-ACCDIST; Brown, 2014), which have been found to be able to assist in automatically recognising speakers' accents for forensic purposes and highlighting key distinguishing variables in different accent groups (Brown & Wormald, 2017). It would therefore be interesting to explore the extent to which MFCCs could capture any variability across speakers from the boroughs of Bradford, Kirklees and Wakefield.

8.5.2. Collect perceptual data

Among others, Pardo et al. (2018b) note that many studies have reported inconsistencies across measures of phonetic convergence in conversational interaction as well as in speech-shadowing tasks. As explained in Section 8.2, only approximately one quarter of the participants in this study accommodated in the same way across FACE and intervocalic /t/ within the same speaking task. This therefore makes it challenging to present an overall view of whether a speaker has truly converged towards or diverged away from their interlocutor, using an exclusively acoustic-phonetic approach. In order to build on this study, it would be useful to collect perceptual similarity ratings in order to capture a more holistic impression of how participants have changed across tasks. A further examination of both acoustic-phonetic and perceptual data would also provide an insight into how noticeable the combination of individual changes within a range of speech parameters are.

Although this investigation has quantified some of the ways in which FACE productions and realisations of intervocalic /t/ have altered both within and between tasks, a perceptual analysis of accommodation was beyond the scope of this study. However, some of the findings in the literature suggest that listeners would be able to perceive the changes observed within many of the participants in this study. For instance, Flanagan (1955) reported that the threshold for just-noticeable differences in vowel perception was approximately 3-5% meaning that changes in F1 or F2 of as little as 3% could be detected by some listeners. Similarly, Labov, Ash, Baranowski, Nagy, & Ravindranath (2006 & 2011) explored listeners' sensitivity to the frequency of sociolinguistic variables and found that their participants showed a "striking consistency in their evaluation of sociolinguistic variables, clearly sensitive to differences in frequency as small as 10%" (2006, p. 127). With regards to varying percentages of the non-standard apical form of the socially salient variable (ING), it was found that listeners had "strong reactions to small percentage differences at the low end of the /ɪn/ scale, and slight reactions to differences at the high end." (2006, p. 127).

Levon & Fox (2014) have suggested that a variable's social salience determines both whether and how it is perceptually evaluated. As speakers appear more likely to accommodate with respect to variables that are socially salient, if listeners are also more sensitive to these variables it seems likely that the accommodation behaviour observed in this study would be perceptible. It has also been found that perceptual awareness is linked to usage (Earnshaw, 2013; Levon & Fox, 2014) and therefore it would be interesting to explore how accommodation in West Yorkshire English is perceived by listeners from inside and outside of the West Yorkshire region. Regarding regional variability, many of the WYRED Task 3 recordings provide a useful resource of meta-commentary about West Yorkshire and how much the accents of Bradford, Kirklees and Wakefield are considered to vary. To supplement this, it would be interesting to conduct a perceptual study in which participants are asked to identify which borough speakers are from, in order to understand the level at which this can be done reliably.

8.5.3. Examine topic shifting effects

A number of studies have demonstrated that changes in topic of conversation can have robust effects on linguistic variation (Devlin, 2014; Douglas-Cowie, 1978; Leach, 2018; Walker, 2014, 2019). For example, Devlin (2014) found that conversational topic was a major constraint on variant usage in East Durham, with the most local variants of GOAT, MOUTH, START and FACE tending to be used most often when speakers were discussing the highly local topic of coal mining. The focus of the present investigation was to observe how speakers accommodate across contexts involving different speaking styles and interlocutors rather than over the course of a conversation involving different topics. However, this would also be a useful and important area to explore further from a forensic perspective, as it is possible to imagine a scenario where the speaking style and interlocutor may be consistent across FSC samples, but topic shifts occur. For instance, the effects of topic shifting may be of relevance in a FSC case involving known and questioned material from the same source where a defence statement of *"everything up to this point was me but someone else was talking from this point onward"* is given.

Using the WYRED data, it would be possible to assess the effects of topic shifts by coding the different topics discussed during the Task 3 casual, paired conversations. Although the participants were given the freedom to discuss any topic they wanted, the same prompt cards were supplied to all participants and therefore most conversations tended to involve at least a subset of the suggested topics. This could lead to a greater understanding of how much speech can vary across different topics within casual paired conversations. It could also be useful to explore whether any particular words or word classes, such as locally iconic words (e.g. *make* and *take*) or proper nouns (e.g. Wakefield), might be influencing any variation in FACE or intervocalic /t/ productions. Due to the semi-spontaneous nature of the speech analysed in this investigation, examples of locally iconic words were unbalanced across speakers and tasks which meant that this could not be reliably examined in a robust way in the present study. However, it would be interesting to explore this by incorporating elicitation of these types of words into the data collection phase of a future investigation.

8.6. Conclusion

In this chapter, answers to all four of the research questions presented in this thesis have been outlined. The differences in FACE and intervocalic /t/ have been highlighted, in terms of local level regional variability and speech accommodation behaviour. An examination of speech accommodation behaviour of all 30 participants included in this study has shown that the consequences of varying speaking styles and interlocutors are highly speaker specific. Overall, the participants' behaviour did not tend to be consistent within features between tasks or within tasks across FACE and /t/. However, the findings indicated that the majority of participants accommodated in their FACE and /t/ productions and the case study of participant #059 showed that socially salient speech parameters were the most variable across tasks. Therefore, the findings of this study are in line with those of previous accommodation studies which have found that speakers tend to accommodate in socially salient variables. The implications of these findings for forensic speech science have been outlined, as have the limitations of the research presented in this thesis and ideas for further areas of study.

9. Conclusion

This chapter concludes the thesis by summarising the research presented, highlighting the key findings of this investigation, and explaining how this research contributes to the fields of forensic speech science and sociophonetics.

9.1. Scope of the research

This thesis has explored regional variation and speech accommodation in two socially salient features of West Yorkshire English. Four research questions were addressed in this project:

1. How much local level variation exists across the West Yorkshire boroughs of Bradford, Kirklees and Wakefield with respect to the FACE vowel and intervocalic /t/?
2. Should reference population data for West Yorkshire speakers be separated by metropolitan borough or is it appropriate to group the boroughs more broadly as "West Yorkshire English"?
3. To what extent do speakers adapt their FACE and intervocalic /t/ productions across forensically-relevant scenarios involving different interlocutors and speaking styles?
4. What are the potential implications of speech accommodation for FSC casework?

In order to address these questions, four case studies were carried out. The speech analysed in each of these studies was from 30 WYRED participants each completing three of the forensically-relevant speaking tasks included in the WYRED database (Gold et al., 2018). The first case study explored how FACE was realised across the West Yorkshire boroughs of Bradford, Kirklees and Wakefield. Levels of regional variability were assessed by measuring midpoint F1~F3 values as well as distance measures from vowel onset to vowel offset for F1 and F2. The second case study examined within-speaker variability in FACE arising from the combined influence of different speaking styles and interlocutors across tasks. A range of methods were used to explore speech accommodation including evaluations of consistency in FACE productions within and between tasks, as well as the calculation of Euclidean distances between interlocutor pairs.

The third case study investigated how word-medial, intervocalic /t/ was realised across West Yorkshire. An auditory analysis of /t/ in three separate phonetic environments revealed the phonetic constraints that apply to this variable in West Yorkshire English and the levels of regional variation across Bradford, Kirklees and Wakefield. Following this, the final case study examined how intervocalic /t/ varied within and between tasks at the group and the individual level. As with the FACE vowel, comparisons of /t/ productions between interlocutor pairs were conducted in order to evaluate the effects of exposure to the interlocutor.

9.2. Key findings

With regards to regional variation within West Yorkshire, it was determined that FACE and intervocalic /t/ patterned in different ways. FACE F2 values varied significantly between the boroughs of Bradford, Kirklees and Wakefield, and therefore appeared to be a marker of local level regional variation. However, realisations of intervocalic /t/ did not vary between the three boroughs. Based on these findings, it was recommended that when establishing typicality of FACE F2 values, the three boroughs should be treated as separate populations, whereas for /t/ or FACE F1 and F3 values, it would be possible to group the boroughs together as "West Yorkshire English" when delimiting the relevant population for FSC casework involving West Yorkshire speakers.

Key findings in relation to speech accommodation in FACE and intervocalic /t/ include the observation that accommodation behaviour appears to be highly speaker specific. With regards to FACE productions, more participants were found to diverge than converge or maintain their original distance from their interlocutor. However, there were examples of all three types of accommodation behaviour for FACE across participants during the mock police interview and the casual paired conversation. It was also observed that more participants displayed maintenance behaviour when interacting with someone from the same area as themselves than when interacting with the female researcher from Gateshead. Overall, FACE realisations tended to be more consistent within each individual during the answer message task where no interlocutor was present compared to during a paired conversation. This indicated that exposure to an interlocutor generally resulted in higher within-speaker variability for FACE.

In relation to word-medial, intervocalic /t/ productions, it was found that the participants accommodated across tasks with respect to their T-glottaling rates. It was evident that most participants tended to have lower T-glottaling rates in the mock police interview task than in the casual paired conversations, in line with the expected style-shifting patterns for a socially salient non-standard variant such as this. It was also observed that the number of participants who converged towards their interlocutor was higher for T-glottaling rates than for FACE productions. Surprisingly, there were more participants who diverged away from their interlocutor over the course of the casual paired conversation than there were in the mock police interview task. It was also found that approximately two thirds of participants displayed different types of accommodation behaviour with respect to T-glottaling rates across the two paired tasks.

The findings of the speech accommodation case studies suggested that experts undertaking FSC casework should exercise particular caution when examining socially salient variables, as it would appear that these variables are most highly influenced by speech accommodation. Although the magnitude of accommodation effects varied across participants, it was found that, for certain participants, realisations of FACE and /t/ varied to such an extent that the within-speaker variability present across samples could potentially be misinterpreted as evidence in support of the different speaker view. It was therefore suggested that a conservative approach should be taken in instances where only a limited range of features could be analysed, especially if these were only socially salient variables. Furthermore, it was recommended that when assessing the suitability of speech samples for FSC casework, experts may find it beneficial to take into account the speech of any interlocutors present in the samples, and consider the potential for speech accommodation to influence the findings.

9.3. Contribution to broader fields

The research presented in this thesis contributes to the growing body of sociophonetic literature in relation to speech accommodation and regional variation. The study provides a new set of population data for FACE and intervocalic /t/ for an area which has received relatively little attention from the sociophonetic community in recent years. This data is of use not only for sociophoneticians but also for forensic phoneticians when making typicality judgements for West Yorkshire English speech. The findings of this research provide an insight into local level regional variation and illustrate that certain speech parameters may be more regionally stratified than previously recognised.

This research also bridges the gap between speech accommodation research and forensic speech science by applying established techniques for measuring accommodation to evaluate within-speaker variability across three forensically-relevant speaking tasks which involve different interlocutors and speaking styles. By exploring the combination of the effects of the interlocutor and speaking style in forensically-relevant speaking situations, this investigation demonstrates the highly speaker specific nature of speech accommodation. Furthermore, in examining semi-spontaneous speech, elicited during face-to-face interactions, the findings of this research are considered to be more applicable to the types of scenarios typically encountered in FSC casework than those of many previous speech accommodation studies which have employed more traditional techniques for data elicitation and analysis. In addition to highlighting the level of potential impact that speech accommodation has on socially salient speech parameters, this investigation opens up an avenue for further exploration of the impact of speech accommodation in FSC casework.

Appendices

Appendix 1: Metadata relating to all 30 participants

Participant ID Number	Age	Borough	Postcode	Years lived outside West Yorkshire	Parents from Bradford, Kirklees or Wakefield?	Employment status	Job role	Phrase identified with the most
#025	19	Bradford	BD11	0	Yes	Student & Part-time employed	Semi or unskilled manual work[5]	British
#030	22	Bradford	BD4	0	Yes	Part-time employed	Semi or unskilled manual work	British
#031	23	Bradford	BD5	0	Yes	Student & Part-time employed	Semi or unskilled manual work	Yorkshire
#033	20	Bradford	BD21	0	Yes	Student & Part-time employed	Casual worker [6]	Yorkshire
#040	24	Bradford	BD17	3	Yes	Student & Part-time employed	Student	Bradford
#042	19	Bradford	BD12	0	Yes	Student	-	Bradford
#064	20	Bradford	BD10	2	No (both from Thirsk)	Student & Part-time employed	Student	Yorkshire
#067	19	Bradford	BD4	0	Yes	Student	-	Yorkshire
#069	20	Bradford	BD18	0	No (Bradford and Leeds)	Student & Part-time employed	Casual worker	British
#022	22	Bradford	BD19	3	No (both from Leeds)	Student	-	West Yorkshire

[5] E.g. Manual workers, apprentices, caretakers, park keepers, Non-HGV drivers, shop assistants
[6] Not in permanent employment

Participant ID Number	Age	Borough	Postcode	Years lived outside West Yorkshire	Parents from Bradford, Kirklees or Wakefield?	Employment status	Job role	Phrase identified with the most
#015	22	Kirklees	HD4	0	Yes	Part-time employed	Semi or unskilled manual work	English
#019	19	Kirklees	HD9	0	No (Glasgow and Hackney)	Student	-	West Yorkshire
#020	26	Kirklees	HD5	6	No (both from Manchester)	Unemployed	-	Huddersfield
#021	24	Kirklees	HD3	1	No (Leeds and Manchester)	Student	-	Huddersfield
#036	21	Kirklees	HD4	0	Yes	Student & Part-time employed	Student	Huddersfield
#046	29	Kirklees	WF17	0	No (Leeds and Liverpool)	Full-time employed	Researcher	British
#048	20	Kirklees	HD2	0	No (both from Sheffield)	Student	-	English
#055	22	Kirklees	WF17	7	No (both from Beverley)	Student	-	Yorkshire
#058	22	Kirklees	WF14	0	Yes	Student	-	Yorkshire
#059	19	Kirklees	HD3	0	Yes	Student & Part-time employed	Casual worker	West Yorkshire

Participant ID Number	Age	Borough	Postcode	Years lived outside West Yorkshire	Parents from Bradford, Kirklees or Wakefield?	Employment status	Job role	Phrase identified with the most
#006	21	Wakefield	WF1	0	Yes	Student & Part-time employed	Semi or unskilled manual work	British
#012	24	Wakefield	WF8	1	No (Wakefield and London)	Student & Part-time employed	Skilled manual worker[7]	English
#034	23	Wakefield	WF4	3	Yes	Student	-	Yorkshire
#035	22	Wakefield	WF4	0	Yes	Student	-	English
#038	24	Wakefield	WF11	0	Yes	Student	-	Yorkshire
#041	19	Wakefield	WF6	0	Yes	Student	-	Wakefield
#045	21	Wakefield	WF3	0	No (both from Hull)	Student & Part-time employed	Supervisory or clerical/ junior managerial / professional / administrative[8]	British
#049	23	Wakefield	WF3	0	Yes	Student	-	British
#050	25	Wakefield	WF4	0	No (Wakefield and Scotland)	Full-time employed	Intermediate managerial/ professional/ administrative[9]	British
#054	20	Wakefield	WF5	0	Yes	Student & Part-time employed	Student	British

[7] E.g. bricklayers, carpenters, plumbers, painters, bus/ambulance drivers, HGV drivers, AA patrolmen, pub/bar workers
[8] E.g. office workers, student doctors, foremen with +25 employees, salespeople
[9] E.g. newly qualified (under 3yrs) doctors, solicitors, board directors in small organisations, middle managers in large organisations, principal officers in the civil service, local government

Appendix 2: Frequency tables for the West Yorkshire participants' FACE and /t/ tokens

FACE Tokens	Frequency	/t/ Tokens	Frequency
say	133	a-forty	108
rachel	105	hotel	99
take	86	pretty	93
same	82	carter	88
day	78	little	68
play	78	peter	56
maybe	74	getting	44
make	73	city	42
straight	58	started	35
steakhouse	55	curtis	32
way	49	vegetarian	32
station	44	theatre	31
place	39	peter's	30
cables	35	scooter	30
eight	35	beetle	29
great	33	eighty	29
wakefield	31	participant	29
name	29	particularly	27
away	28	forty	25
mate	28	better	24
days	27	pighty	24
came	26	typesetter	24
capers	25	certain	21
eighty	25	thirty	21
change	21	community	15
steak	21	important	15
made	20	whatever	15
taken	20	later	12
main	19	quarter	12
bakers	17	notice	10
places	17	noticed	10
stay	17	bottom	9
takes	17	charity	9
saying	16	fourteen	8
mates	15	sorted	8
playing	14	thirteen	8
baker's	13	university	8
may	13	computers	7
placement	13	guitar	7
played	13	politics	7
game	12	tutor	7
hate	12	beautiful	6
mainly	12	particular	6

FACE Tokens	Frequency	/t/ Tokens	Frequency
faces	11	united	6
games	11	chatting	5
names	11	creative	5
pay	11	eighteen	5
safe	11	hated	5
save	11	incriminating	5
makes	10	positive	5
paid	10	scottish	5
strange	10	automatic	4
based	9	beetles	4
late	9	british	4
face	8	chatted	4
making	8	exciting	4
chain	7	footage	4
later	7	located	4
plate	7	mutated	4
rachel's	7	political	4
taking	7	putting	4
break	6	saturday	4
paper	6	shooting	4
stayed	6	sitting	4
training	6	skating	4
wait	6	solicitor	4
basis	5	visiting	4
cable	5	water	4
changed	5	cater	3
date	5	computer	3
facebook	5	detained	3
gave	5	eating	3
pain	5	forgotten	3
wakey	5	graduated	3
afraid	4	invited	3
case	4	metal	3
explain	4	monitored	3
hated	4	motorway	3
name's	4	party	3
safety	4	recruiting	3
sake	4	retail	3
space	4	stereotypical	3
train	4	totally	3
ways	4	waterproof	3
age	3	articles	2
aim	3	attached	2
baker	3	attack	2
brain	3	attention	2

FACE Tokens	Frequency	/t/ Tokens	Frequency
complain	3	bartender	2
crazy	3	britain	2
detained	3	butty	2
estate	3	committed	2
grades	3	communities	2
jail	3	competition	2
mistake	3	competitive	2
pace	3	computing	2
paint	3	duty	2
pale	3	fitting	2
papers	3	idiosyncratic	2
phrases	3	italy	2
plays	3	majority	2
rain	3	manhattan	2
saves	3	matter	2
saving	3	meeting	2
state	3	monotone	2
steakhouses	3	motorbike	2
table	3	neverton	2
taste	3	opportunities	2
ashamed	2	opportunity	2
bacon	2	reputation	2
bakery	2	saturdays	2
basic	2	scattered	2
blame	2	scooters	2
cater	2	soliciting	2
chains	2	solicitors	2
changing	2	solicitor's	2
craig	2	sporting	2
dates	2	statistics	2
delay	2	touting	2
faith	2	unfacilitated	2
grange	2	universities	2
james	2	visited	2
lazy	2	voters	2
major	2	waited	2
named	2	waiting	2
paying	2	writing	2
players	2	activator	1
rach	2	activities	1
rage	2	activity	1
relate	2	advertise	1
replace	2	advertisement	1
replaced	2	advertising	1
says	2	alliteration	1

FACE Tokens	Frequency	/t/ Tokens	Frequency
scale	2	allocated	1
shaky	2	analytical	1
shape	2	anxiety	1
skatepark	2	appreciated	1
skating	2	architect	1
spaces	2	architecture	1
spaceship	2	articulate	1
spain	2	artificial	1
staying	2	asserted	1
that	2	attitude	1
today	2	auto	1
traces	2	bartenders	1
waveforms	2	battered	1
able	1	battering	1
acres	1	betting	1
against	1	bottle	1
ages	1	brutal	1
aid's	1	butter	1
ale	1	capital	1
ape	1	categories	1
april	1	charity's	1
asian	1	chatty	1
ate	1	christianity	1
bail	1	cities	1
bases	1	committing	1
basics	1	commuting	1
bassist	1	complicated	1
became	1	creator	1
boy	1	database	1
brakes	1	daughter	1
breakdown	1	dedicated	1
breaking	1	deleting	1
breaks	1	delighted	1
cake	1	departing	1
cape	1	detail	1
cases	1	details	1
casey	1	detention	1
claimed	1	detentions	1
claiming	1	determined	1
claims	1	determines	1
crane	1	dissertation	1
crates	1	distributed	1
daily	1	diverted	1
damon	1	dominated	1
danger	1	donated	1

FACE Tokens	Frequency	/t/ Tokens	Frequency
day's	1	dramatically	1
decay	1	duties	1
delayed	1	editing	1
disgrace	1	excited	1
drain	1	exotic	1
eighteen	1	fanatic	1
engage	1	fifty	1
essays	1	fighters	1
failed	1	footy	1
fame	1	glottal	1
framework	1	greatest	1
gaping	1	guatemala	1
gay	1	heated	1
grade	1	hotels	1
grateful	1	hotel's	1
greatly	1	interpretation	1
haiches	1	invigilator	1
hates	1	kiting	1
jane	1	lambretta	1
ladies	1	lighter	1
lake	1	limited	1
lane	1	literal	1
lay	1	marketing	1
lay-by	1	material	1
mainstream	1	materials	1
male	1	mediterranean	1
mate's	1	mentality	1
mutate	1	millimetre	1
naming	1	minnesota	1
obey	1	monotony	1
pacing	1	motivation	1
page	1	mutate	1
painted	1	native	1
paints	1	nominated	1
payments	1	parties	1
payphone	1	peters	1
payphones	1	petted	1
persuade	1	petty	1
phases	1	pigity	1
phrase	1	plighty	1
placements	1	possibility	1
plain	1	potentially	1
plates	1	potter	1
player	1	potters	1
racer	1	pretend	1

FACE Tokens	Frequency	/t/ Tokens	Frequency
racist	1	pretenders	1
raining	1	prioritise	1
raping	1	priority	1
rapist	1	proprietor	1
rates	1	qualitative	1
rein	1	quality	1
remains	1	quantitative	1
safer	1	quieter	1
saint	1	rated	1
saved	1	ratty	1
saver	1	recruiters	1
scraped	1	regretting	1
scraping	1	relatively	1
shaking	1	repeating	1
shaky	1	reporter	1
shame	1	retails	1
shameless	1	routine	1
shapes	1	security	1
skate	1	sensitive	1
stable	1	settee	1
stage	1	setting	1
states	1	settled	1
stations	1	shouting	1
station's	1	shutting	1
steaks	1	situated	1
strangers	1	society	1
tastes	1	sorting	1
trace	1	stability	1
trade	1	supporters	1
tradesmen	1	supporting	1
trainers	1	sweaty	1
trains	1	title	1
upgrade	1	trinity	1
wage	1	tutors	1
waiter	1	twitter	1
waiting	1	undateables	1
waitress	1	voting	1
wake	1	waiter	1
wasted	1	waterproofs	1
wave	1	waterway	1
waving	1	witty	1
TOTAL	**2119**		**1599**

Appendix 3: FACE linear mixed effects model summaries

FACE F1 distance measures with by-participant random slopes for TASK (failed to converge)

	Estimate	Std. Error	df	t value
Intercept (Bradford, Task 4, main environment group)	-13.051	10.031	39	-1.30
Kirklees	-1.856	12.633	30	-0.15
Wakefield	-26.985	12.597	30	-2.14
Task 1	2.899	4.740	33	0.61
Task 3	0.728	4.202	29	0.17
Environment 1: liquid__nasal	15.201	7.269	2069	2.09
Environment 2: __nasal	25.929	3.393	2079	7.64
Environment 3: glide__	-10.091	4.607	2075	-2.19
Environment 4: __liquid	23.791	14.565	2025	1.63
Environment 5: liquid __	-7.430	2.820	2083	-2.64
Environment 6: *make* and *take*	-21.032	4.031	2080	-5.22
Environment 7: <eigh> spelling	-91.896	6.546	2049	-14.04

FACE F1 distance measures without by-participant random slopes for TASK (converged)

	Estimate	Std. Error	df	t value
Intercept (Bradford, Task 4, main environment group)	-13.614	9.530	38	-1.43
Kirklees	-0.280	12.676	30	-0.02
Wakefield	-26.188	12.659	30	-2.07
Task 1	2.960	3.518	2090	0.84
Task 3	0.796	3.548	2089	0.22
Environment 1: liquid__nasal	14.322	7.307	2089	1.96
Environment 2: __nasal	25.267	3.406	2089	7.42
Environment 3: glide__	-10.160	4.627	2088	-2.20
Environment 4: __liquid	24.205	14.533	2090	1.67
Environment 5: liquid __	-7.738	2.826	2088	-2.74
Environment 6: *make* and *take*	-22.060	4.038	2088	-5.46
Environment 7: <eigh> spelling	-92.524	6.584	2087	-14.05

FACE F2 distance measures

	Estimate	Std. Error	df	t value
Intercept (Bradford, Task 4, main environment group)	54.272	18.321	43	2.96
Kirklees	-5.354	21.577	30	-0.25
Wakefield	33.066	21.530	30	1.54
Task 1	-7.831	11.413	32	-0.69
Task 3	5.490	10.419	27	0.53
Environment 1: liquid__nasal	117.508	16.506	2065	7.12
Environment 2: __nasal	-54.506	7.699	2076	-7.08
Environment 3: glide__	178.336	10.463	2074	17.05
Environment 4: __liquid	-11.177	33.142	2053	-0.34
Environment 5: liquid __	157.577	6.404	2085	24.61
Environment 6: *make* and *take*	66.385	9.148	2079	7.26
Environment 7: <eigh> spelling	151.884	14.836	2044	10.24

FACE F1 midpoint values with by-participant random slopes for TASK (failed to converge)

	Estimate	Std. Error	df	t value
Intercept (Bradford, Task 4, main environment group)	552.608	8.555	31	64.60
Kirklees	-39.115	11.764	28	-3.33
Wakefield	-12.729	11.772	28	-1.08
Task 1	10.255	3.550	71	2.89
Task 3	4.166	3.915	27	1.06
Environment 1: liquid__nasal	13.619	6.628	2080	2.06
Environment 2: __nasal	20.707	3.086	2070	6.71
Environment 3: glide__	8.740	4.198	2079	2.08
Environment 4: __liquid	28.352	13.192	2046	2.15
Environment 5: liquid __	4.274	2.562	2074	1.67
Environment 6: *make* and *take*	6.850	3.655	2064	1.87
Environment 7: <eigh> spelling	-27.463	5.955	2065	-4.61

FACE F1 midpoint values without by-participant random slopes for TASK (converged)

	Estimate	Std. Error	df	t value
Intercept (Bradford, Task 4, main environment group)	548.009	9.696	36	56.52
Kirklees	-30.865	13.059	30	-2.36
Wakefield	-6.394	13.045	30	-0.49
Task 1	10.024	3.212	2089	3.12
Task 3	3.519	3.239	2088	1.09
Environment 1: liquid__nasal	14.548	6.671	2088	2.18
Environment 2: __nasal	20.490	3.110	2088	6.59
Environment 3: glide__	9.841	4.224	2087	2.33
Environment 4: __liquid	28.218	13.270	2089	2.13
Environment 5: liquid __	4.427	2.580	2088	1.72
Environment 6: *make* and *take*	7.263	3.687	2088	1.97
Environment 7: <eigh> spelling	-26.524	6.011	2087	-4.41

FACE F2 midpoint values

	Estimate	Std. Error	df	t value
Intercept (Bradford, Task 4, main environment group)	1810.051	25.587	35	70.74
Kirklees	-25.753	33.212	29	-0.78
Wakefield	90.031	33.194	29	2.71
Task 1	-56.689	14.063	31	-4.03
Task 3	-59.574	13.448	29	-4.43
Environment 1: liquid__nasal	-48.646	17.422	2061	-2.79
Environment 2: __nasal	6.499	8.135	2066	0.80
Environment 3: glide__	-106.115	11.050	2068	-9.60
Environment 4: __liquid	-46.216	35.116	2078	-1.32
Environment 5: liquid __	-151.018	6.773	2080	-22.30
Environment 6: *make* and *take*	-5.533	9.671	2072	-0.57
Environment 7: <eigh> spelling	205.086	15.651	2042	13.10

FACE F3 midpoint values

	Estimate	Std. Error	df	t value
Intercept (Bradford, Task 4, main environment group)	2520.603	33.673	31	74.86
Kirklees	11.568	46.592	30	0.25
Wakefield	86.606	46.608	30	1.86
Task 1	14.122	11.386	30	1.24
Task 3	-2.832	9.309	27	-0.30
Environment 1: liquid__nasal	-120.691	17.400	2059	-6.94
Environment 2: __nasal	-27.664	8.104	2071	-3.41
Environment 3: glide__	-161.888	11.021	2070	-14.69
Environment 4: __liquid	-74.939	34.750	1977	-2.16
Environment 5: liquid __	-119.930	6.736	2074	-17.80
Environment 6: *make* and *take*	-24.221	9.616	2067	-2.52
Environment 7: <eigh> spelling	68.449	15.632	2038	4.38

FACE F1 SD values

	Estimate	Std. Error	df	t value
Intercept (Bradford, Task 4)	39.689	3.319	53	11.96
Kirklees	-5.759	3.991	30	-1.44
Wakefield	0.111	3.991	30	0.03
Task 1	5.133	3.025	60	1.70
Task 3	6.696	3.025	60	2.21

FACE F2 SD values

	Estimate	Std. Error	df	t value
Intercept (Bradford, Task 4)	99.629	8.549	44	11.65
Kirklees	-5.878	10.921	30	-0.54
Wakefield	-1.077	10.921	30	-0.10
Task 1	39.808	6.352	60	6.27
Task 3	44.557	6.352	60	7.02

FACE F3 SD values

	Estimate	Std. Error	df	t value
Intercept (Bradford, Task 4)	131.815	11.126	53	11.85
Kirklees	-22.789	13.442	30	-1.70
Wakefield	-19.240	13.442	30	-1.43
Task 1	4.623	10.017	60	0.46
Task 3	9.947	10.017	60	0.99

FACE *difference in distance* values for both tasks

	Estimate	Std. Error	df	t value
Intercept (Bradford, Task 1, main environment group)	-0.055	0.048	42	-1.14
Kirklees	0.149	0.060	30	2.50
Wakefield	0.053	0.059	28	0.89
Task 3	-0.045	0.027	29	-1.65
Environment 1: liquid__nasal	0.075	0.064	1784	1.18
Environment 2: __nasal	0.036	0.031	1779	1.18
Environment 3: glide__	0.105	0.042	1784	2.49
Environment 4: __liquid	-0.074	0.153	1778	-0.49
Environment 5: liquid __	0.289	0.027	1782	10.78
Environment 6: *make* and *take*	0.053	0.040	1774	1.34
Environment 7: <eigh> spelling	0.087	0.057	1763	1.52

FACE *difference in distance* values for Task 1

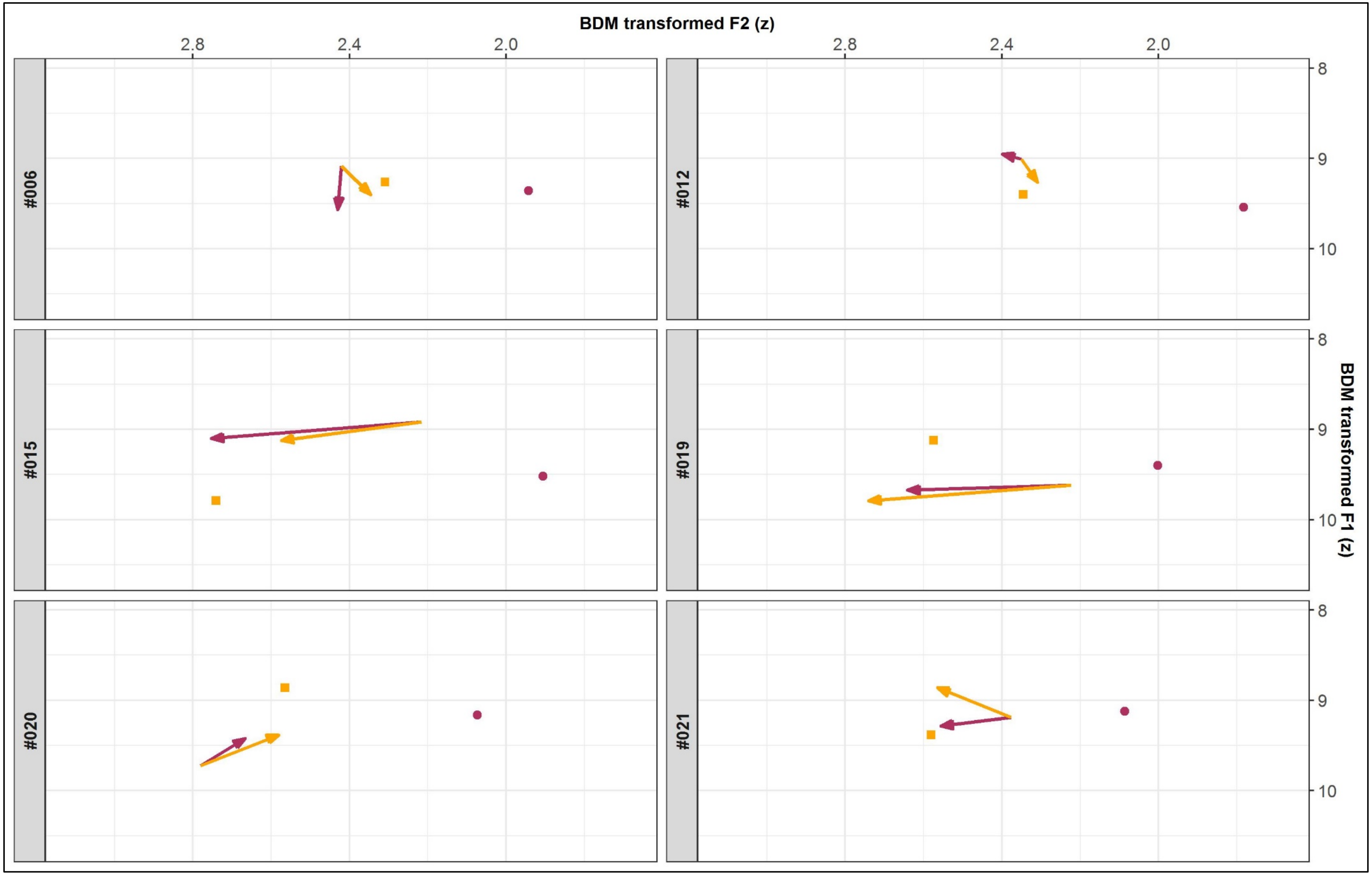
	Estimate	Std. Error	df	t value
Intercept (Bradford, Early, main environment group)	-0.084	0.061	43	-1.38
Kirklees	0.208	0.078	30	2.66
Wakefield	0.119	0.078	30	1.52
Late	-0.018	0.030	918	-0.62
Environment 1: liquid__nasal	0.262	0.113	923	2.31
Environment 2: __nasal	-0.035	0.044	925	-0.79
Environment 3: glide__	0.133	0.073	924	1.82
Environment 4: __liquid	-0.141	0.296	924	-0.48
Environment 5: liquid __	0.313	0.038	928	8.24
Environment 6: *make* and *take*	0.006	0.053	925	0.11
Environment 7: <eigh> spelling	0.075	0.062	921	1.21

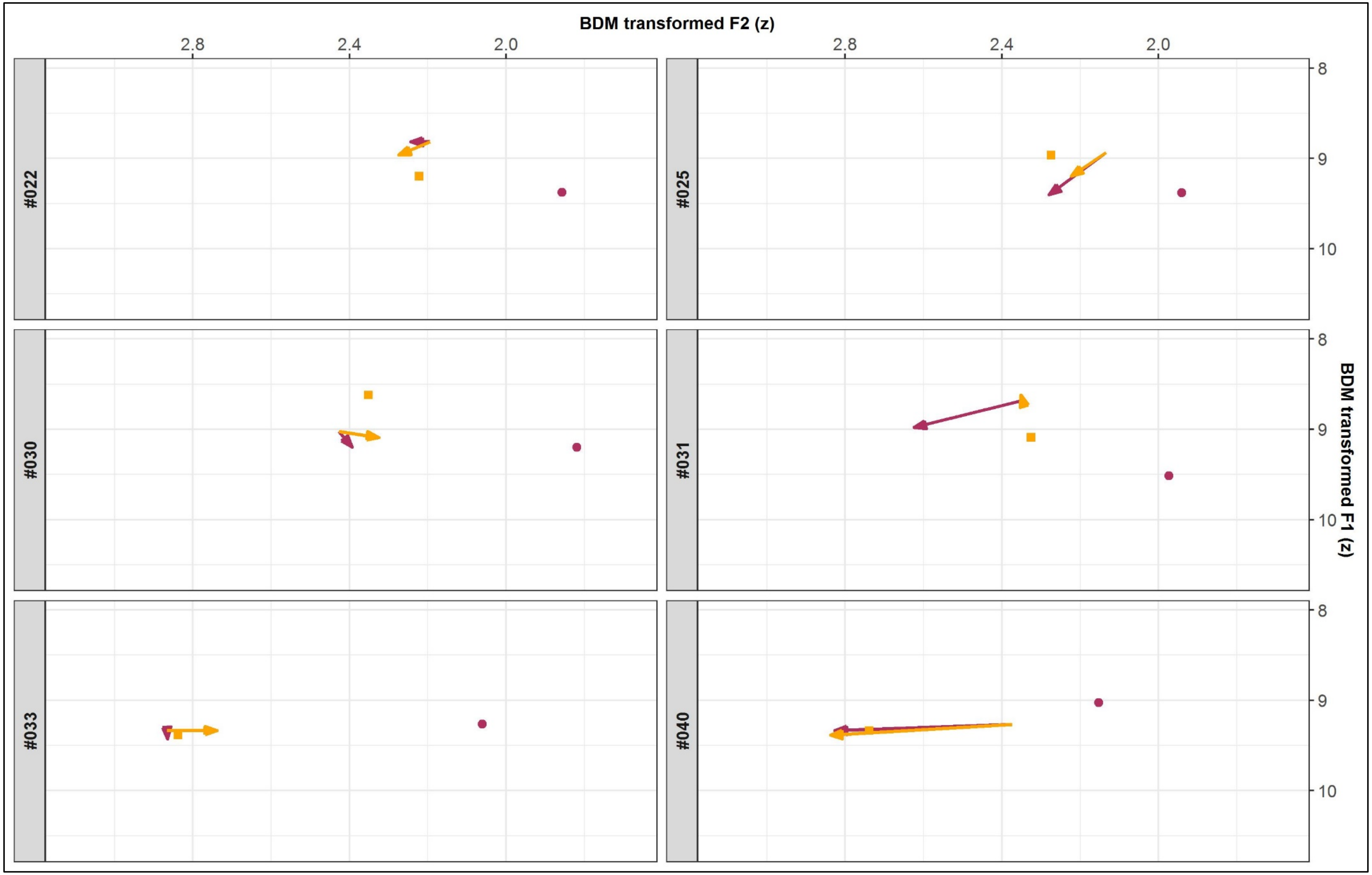
FACE *difference in distance* values for Task 3

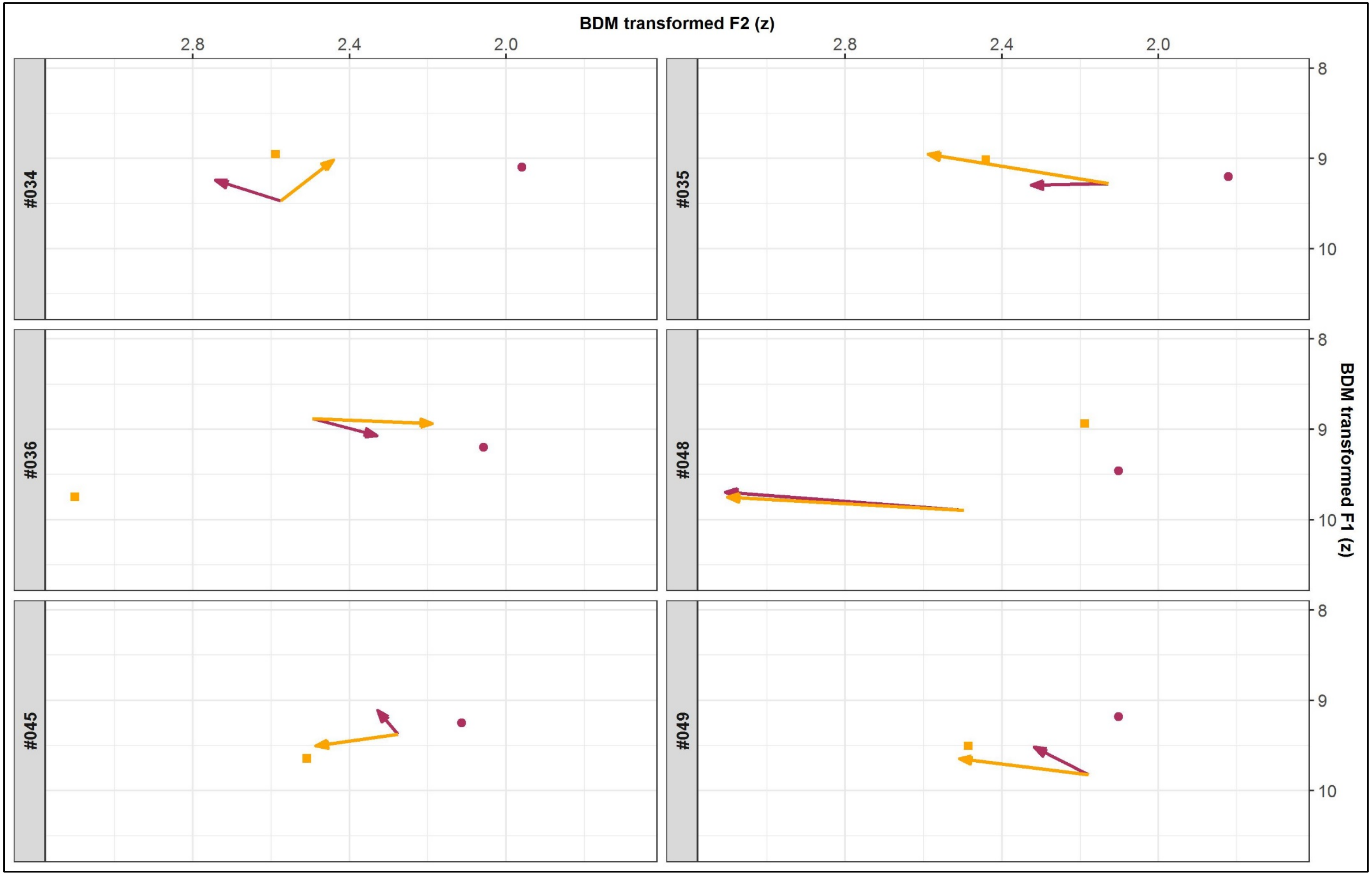
	Estimate	Std. Error	df	t value
Intercept (Bradford, Early, main environment group)	-0.075	0.048	48	-1.55
Kirklees	0.126	0.061	30	2.07
Wakefield	0.038	0.059	28	0.65
Late	-0.029	0.030	834	-0.98
Environment 1: liquid__nasal	0.001	0.079	852	0.01
Environment 2: __nasal	0.085	0.043	849	1.97
Environment 3: glide__	0.079	0.053	849	1.49
Environment 4: __liquid	-0.023	0.182	862	-0.13
Environment 5: liquid __	0.256	0.040	846	6.38
Environment 6: *make* and *take*	0.100	0.061	842	1.64
Environment 7: <eigh> spelling	0.125	0.180	851	0.70

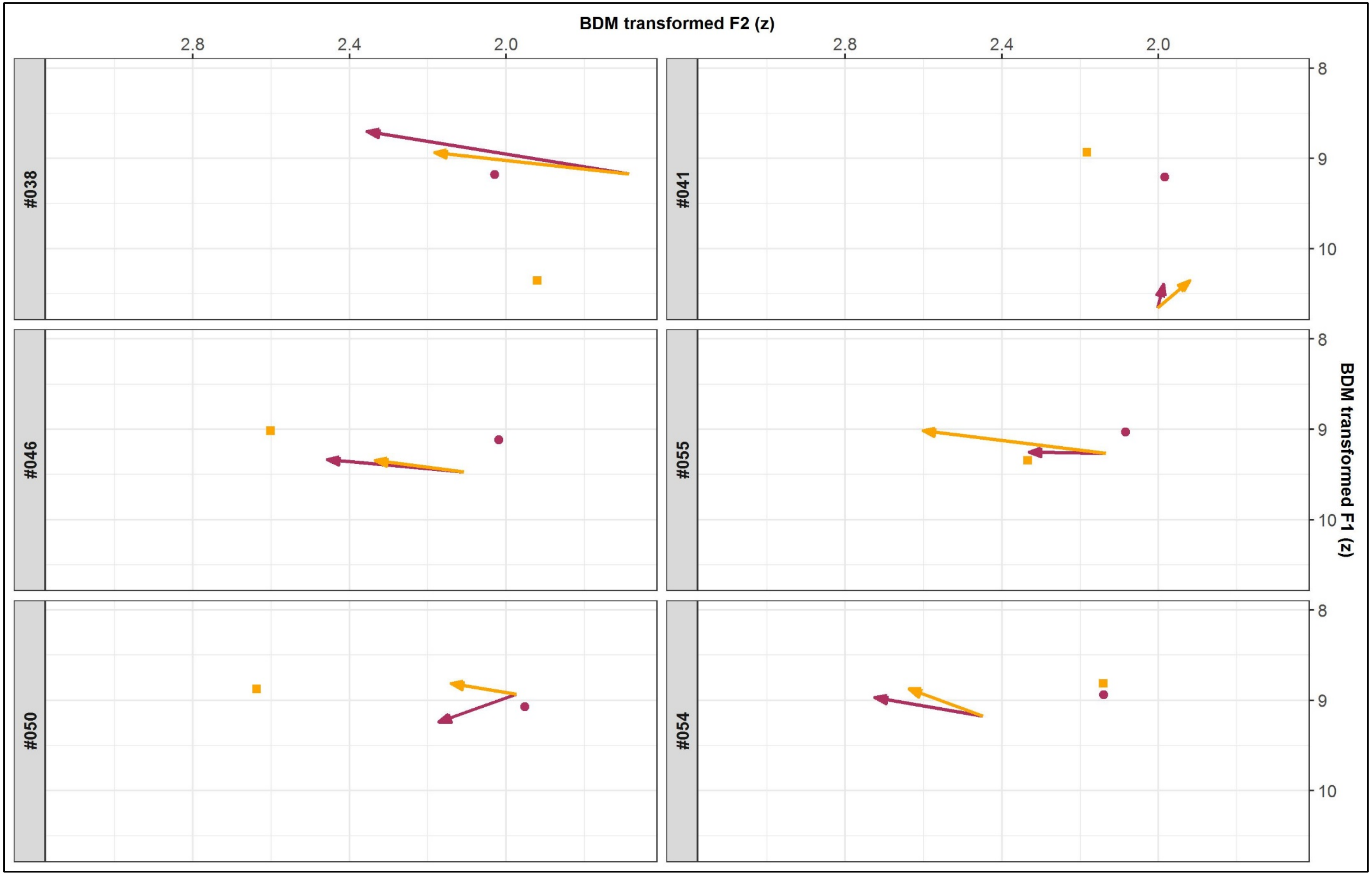
Appendix 4: FACE accommodation plots for all 30 participants

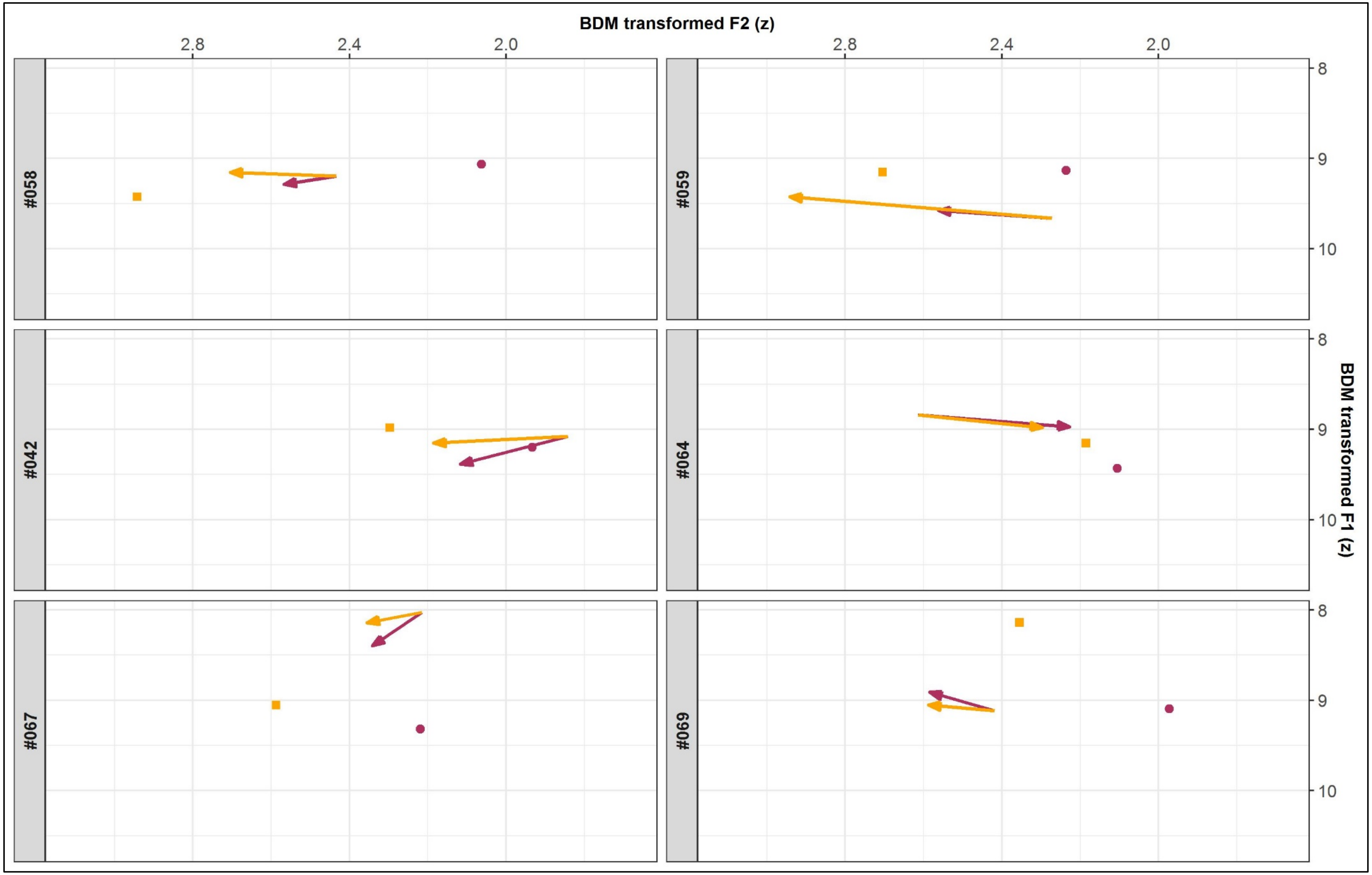
The summary plots for each of the participants presented below represent the participants' average BDM transformed FACE productions across tasks with their respective interlocutors' FACE productions marked for reference. Participants who were paired together during the Task 3 recordings are presented on the same row (e.g. participant #006 and #012 were partners). In each plot, the starting point of the two arrows represents the vowel in the Task 4 baseline task (average transformed F1 and F2), while the end point represents the vowel in the paired task (average transformed F1 and F2). The maroon arrow represents the difference between the baseline and Task 1, while the yellow arrow represents the difference between the baseline and Task 3. The maroon circle represents the Task 1 interlocutor's average transformed FACE vowel and the yellow square represents the average transformed FACE vowel of the Task 3 interlocutor. Please note that the x- and y-axes are reversed in line with standard conventions, in order to better represent the vowel space.


BDM transformed F2 (z)
2.8
2.4
2.0
BDM transformed F1 (z)
8
9
10
#006
#012
#015
#019
#020
#021


BDM transformed F2 (z)
2.8
2.4
2.0
BDM transformed F1 (z)
8
9
10
#022
#025
#030
#031
#033
#040

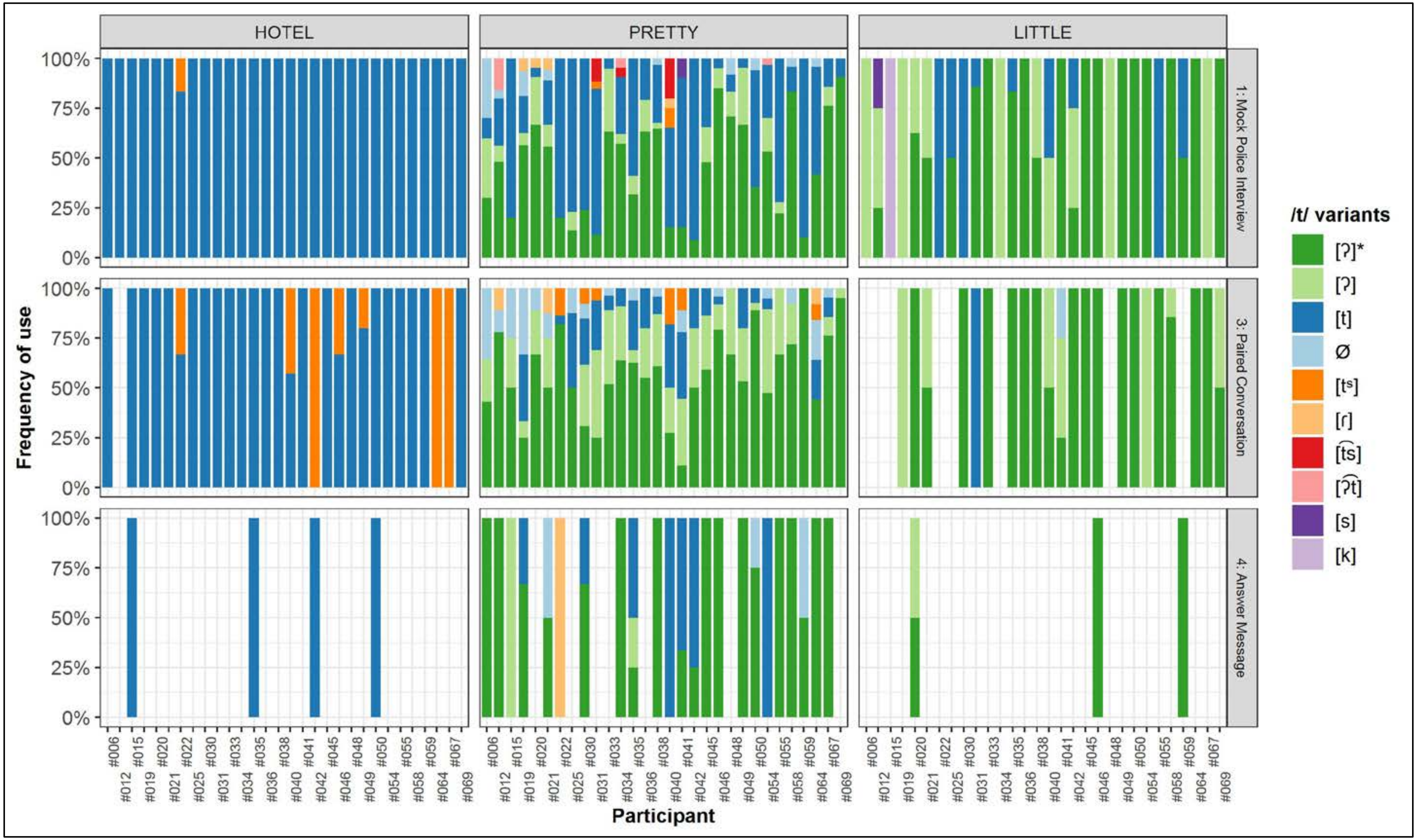

BDM transformed F2 (z)
2.8
2.4
2.0
BDM transformed F1 (z)
8
9
10
#034
#035
#036
#048
#045
#049


BDM transformed F2 (z)
2.8
2.4
2.0
BDM transformed F1 (z)
8
9
10
#038
#041
#046
#055
#050
#054



BDM transformed F2 (z)
2.8
2.4
2.0
BDM transformed F1 (z)
8
9
10
#058
#059
#042
#064
#067
#069

Appendix 5: /t/ distribution plots for all 30 participants by phonetic environment and task



Appendix 6: /t/ distribution plots for all 30 participants and their respective interlocutors in Task 1 and 3 by phonetic environment


HOTEL
PRETTY
LITTLE
Frequency of use
100%
75%
50%
25%
0%
#022
#025
#030
#031
#033
#040
T1 Participant
T1 Interlocutor
T3 Participant
T3 Interlocutor
Task and Speaker
/t/ variants
[ʔ]*
[ʔ]
[t]
Ø
[tˢ]
[ɾ]
[t͡s]
[ʔ͡t]
[s]
[k]
[ʔ͡ɾ]

HOTEL
PRETTY
LITTLE
100%
75%
50%
25%
0%
Frequency of use
#034
#035
#036
#048
#045
#049
T1 Participant
T1 Interlocutor
T3 Participant
T3 Interlocutor
Task and Speaker
/t/ variants
[ʔ]*
[ʔ]
[t]
Ø
[tˢ]
[ɾ]
[t͡s]
[ʔ͡t]
[s]
[k]
[ʔ͡ɾ]


HOTEL
PRETTY
LITTLE
100%
75%
50%
25%
0%
Frequency of use
#038
#041
#046
#055
#050
#054
T1 Participant
T1 Interlocutor
T3 Participant
T3 Interlocutor
Task and Speaker
/t/ variants
[ʔ]*
[ʔ]
[t]
Ø
[tˢ]
[ɾ]
[t͡s]
[ʔ͡t]
[s]
[k]
[ʔ͡ɾ]

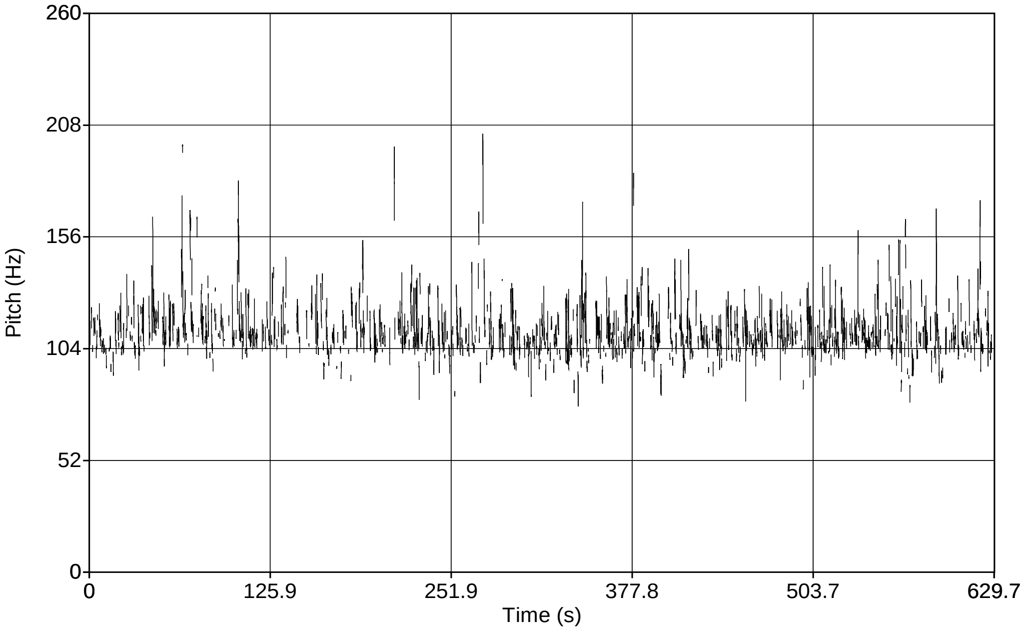

HOTEL
PRETTY
LITTLE
100%
75%
50%
25%
0%
Frequency of use
#058
#059
#042
#064
#067
#069
T1 Participant
T1 Interlocutor
T3 Participant
T3 Interlocutor
Task and Speaker
/t/ variants
[ʔ]*
[ʔ]
[t]
ø
[tˢ]
[ɾ]
[t͡s]
[ʔ͡t]
[s]
[k]
[ʔ͡ɾ]

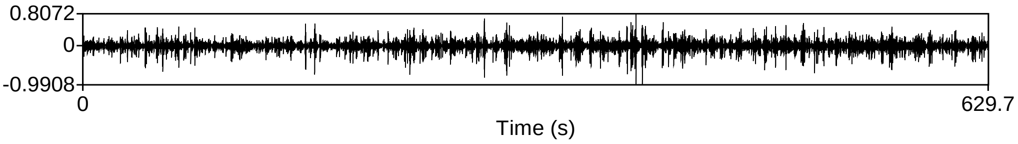
Appendix 7: Details of participant #059's f_0 analysis results and settings used for analysis

Task 1: Mock Police Interview

Case: Participant **0**59 - 059
Speaker: **0**59
Source of sample: Task 1 - F0
Analyst: Katherine Earnshaw

Sample duration = 630 s
Extracted speech
Questioned sample
Time & Date: 11:39:52 Sat 07 Nov 2020





Analysis Settings

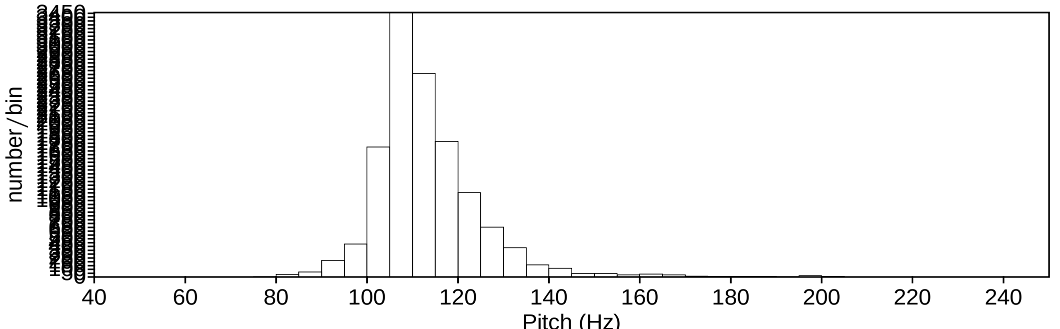
Time step = 0.01 s
Minimum pitch = 75 Hz
Max. number of candidates = 15
Very accurate = no
Silence threshold = 0.05
Voicing threshold = 0.45
Octave cost = 0.01
Octave jump cost = 2.5
Voiced/unvoiced cost = 0.5
Maximum pitch = 250 Hz
Filename: D: D 9 Audio & Transcripts 9-1-K-S-14022017 - F0 edit n.wav

Analysis Results

F_0 = 113 Hz
Standard Deviation = 12 Hz
Minimum = 77 Hz
Maximum = 204 Hz
Median = 111 Hz
Alternative Baseline = 101 Hz

Pitch Distribution

Bin width = 5 Hz
Pitch distribution logged
File: 059-1-K-S-14022017 - F0 edit n - Pitch10.dis

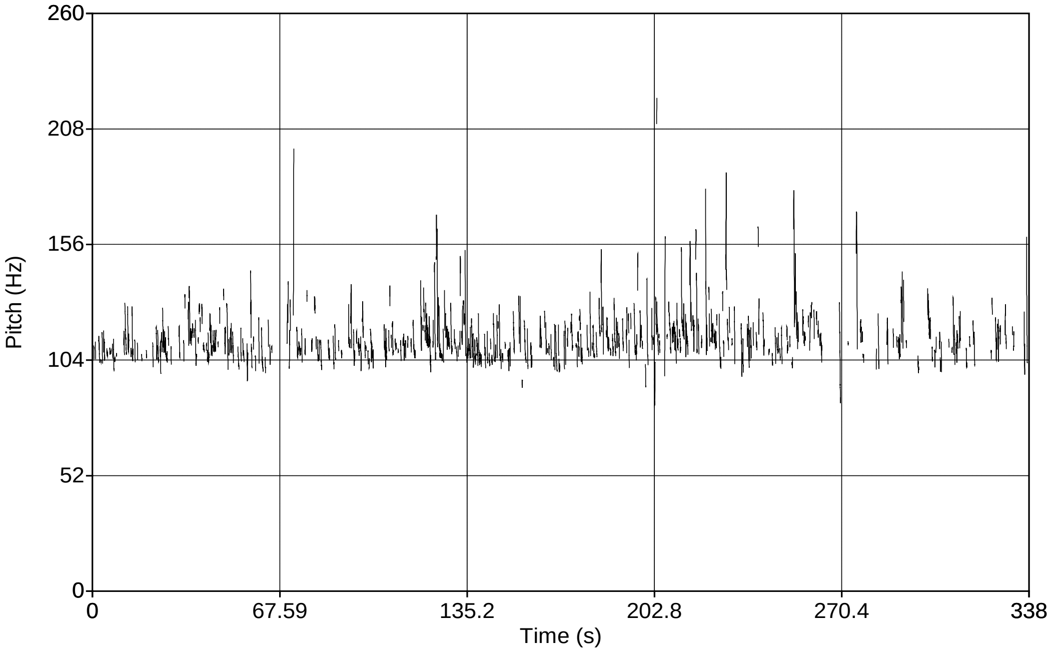


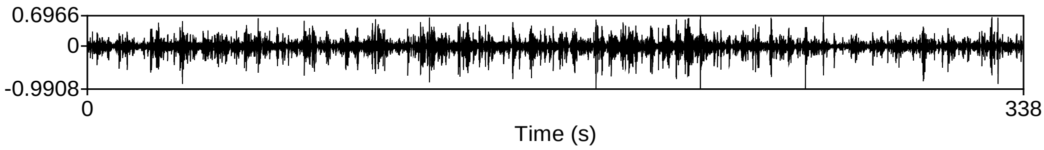
Computer: ELITEBOOK-NO2
Praat Version: 6.1.09

Username: Kate Earnshaw
Script Name: f0Analysis.praat - Version: 1.9.4

Task 3: Paired conversation

Case: Participant **0**59 - 059 | Sample duration = 338 s
Speaker: **0**59 | Extracted speech
Source of sample: Task 3 - F0 | Questioned sample
Analyst: Katherine Earnshaw | Time & Date: 17:04:37 Sat 07 Nov 2020





Analysis Settings

Time step = 0.01 s
Minimum pitch = 75 Hz
Max. number of candidates = 15
Very accurate = no
Silence threshold = 0.05
Voicing threshold = 0.3
Octave cost = 0.2
Octave jump cost = 4
Voiced/unvoiced cost = 1
Maximum pitch = 250 Hz

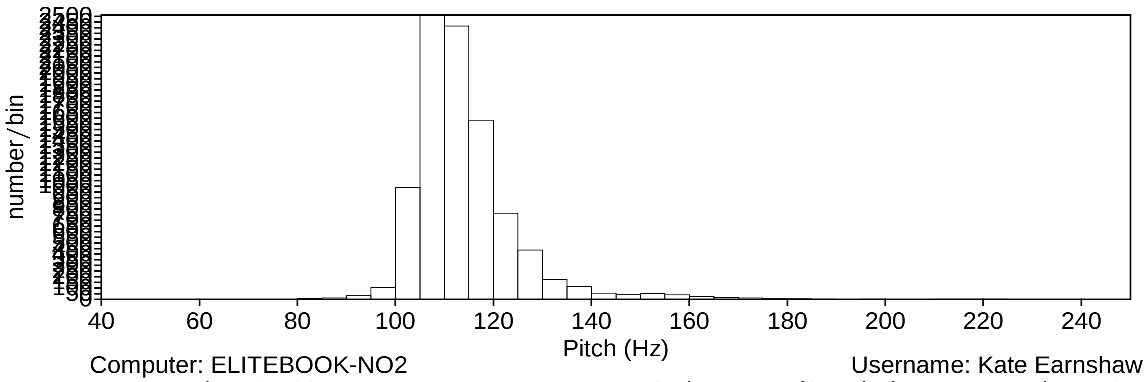
Analysis Results

F_0 = 114 Hz
Standard Deviation = 11 Hz
Minimum = 83 Hz
Maximum = 222 Hz
Median = 112 Hz
Alternative Baseline = 104 Hz

Filename: D: D 9 Audio & Transcripts 9-3-K-S-NF-058-27022017 - F0 edit n.wav

Pitch Distribution

Bin width = 5 Hz
Pitch distribution logged
File: 059-3-K-S-NF-058-27022017 - F0 edit n - Pitch7.dis

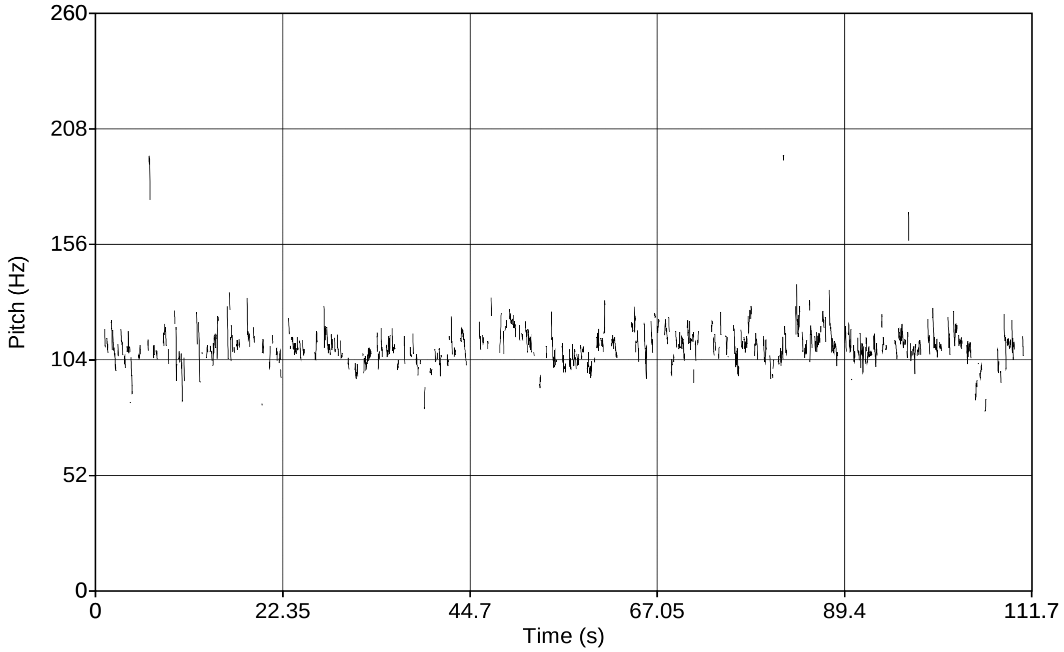


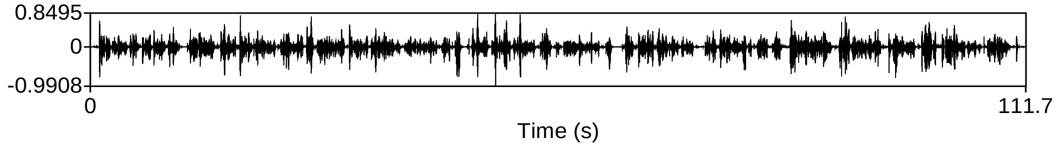
Computer: ELITEBOOK-NO2 | Username: Kate Earnshaw
Praat Version: 6.1.09 | Script Name: f0Analysis.praat - Version: 1.9.4

Task 4: Answer Message

Case: Participant **0**59 - 059
Speaker: **0**59
Source of sample: Task 4 - Answer Message
Analyst: Katherine Earnshaw

Sample duration = 112 s
Extracted speech
Known sample
Time & Date: 17:12:05 Sat 07 Nov 2020





Analysis Settings

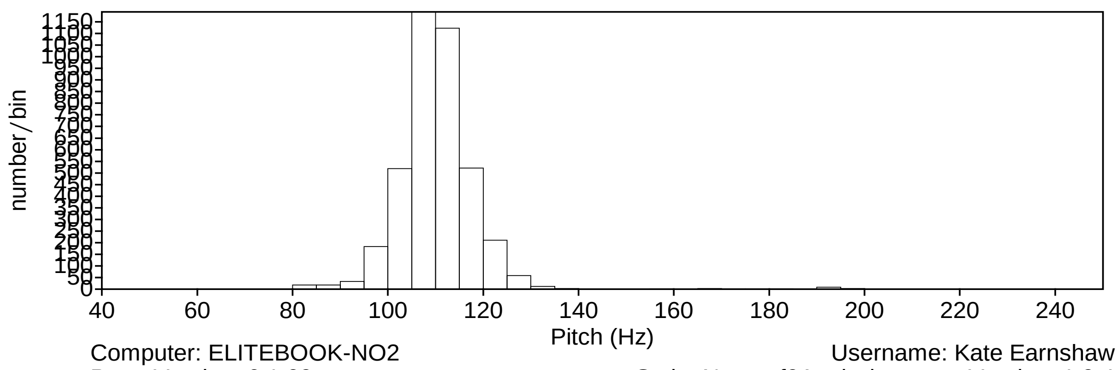
Time step = 0.01 s
Minimum pitch = 75 Hz
Max. number of candidates = 15
Very accurate = no
Silence threshold = 0.03
Voicing threshold = 0.45
Octave cost = 0.1
Octave jump cost = 3
Voiced/unvoiced cost = 0.25
Maximum pitch = 250 Hz
Filename: D: D 9 Audio & Transcripts 9-4-K-S-27022017- edit n.wav

Analysis Results

F_0 = 110 Hz
Standard Deviation = 9 Hz
Minimum = 81 Hz
Maximum = 196 Hz
Median = 110 Hz
Alternative Baseline = 101 Hz

Pitch Distribution

Bin width = 5 Hz
Pitch distribution logged
File: 059-4-K-S-27022017- edit n - Pitch9.dis



Computer: ELITEBOOK-NO2
Praat Version: 6.1.09

Username: Kate Earnshaw
Script Name: f0Analysis.praat - Version: 1.9.4

References

Aguilar, L., Downey, G., Krauss, R. M., Pardo, J. S., Lane, S., & Bolger, N. (2016). A Dyadic Perspective on Speech Accommodation and Social Connection: Both Partners' Rejection Sensitivity Matters. *Journal of Personality*, 84(2), 165–177.

Alderton, R. (2020). Perceptions of T-glottalling among adolescents in South East England. *English Today*, 36(3), 40–47.

Alexander, A., Forth, O., Atreya, A. A., & Kelly, F. (2016). VOCALISE : A forensic automatic speaker recognition system supporting spectral, phonetic, and user-provided features. In *Odyssey*.

Alshangiti, W., & Evans, B. G. (2011). Regional Accent Accommodation in Spontaneous Speech: Evidence for Long-Term Accent Change? In *Proceedings of ICPhS 17, Hong Kong* (pp. 224–227).

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., … Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34(4), 351–366.

Assmann, P. (2008). Developmental study of the relationship between F0 and formant frequencies. *The Journal of the Acoustical Society of America*, 124(4), 2556–2556.

Auer, P., & Hinskens, F. (2005). The role of interpersonal accommodation in a theory of language change. In P. Auer, F. Hinskens, & P. Kerswill (Eds.), *Dialect Change: Convergence and Divergence in European Languages* (pp. 335–357). Cambridge: Cambridge University Press.

Babel, M. (2009). *Phonetic and Social Selectivity in Speech Accommodation*. (Doctoral thesis). University of California, Berkley.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, 39(4), 437–456.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), 177–189.

Babel, M., & Bulatov, D. (2012). The Role of Fundamental Frequency in Phonetic Accommodation. *Language and Speech*, 55(2), 231–248.

Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, 5(1), 123–150.

Bailly, G., & Martin, A. (2014). Assessing objective characterizations of phonetic convergence.

In *Proceedings of Interspeech 2014, Singapore* (pp. 19–9).

Baker, R., & Hazan, V. (2011). DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, 43(3), 761–770.

Baranowski, M., & Turton, D. (2015). Manchester English. In R. Hickey (Ed.), *Researching Northern Englishes* (pp. 293–316). Amsterdam and Philadelphia: John Benjamins Publishing Company.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Beeman, W. O. (2005). *Persian, Dari and Tajik in Central Asia*. *The National Council for Eurasian and East European Research*.

Bell, A. (1984). Language style as audience design. *Language in Society*, 13(2), 145–204.

Beňuš, Š., Gravano, A., & Hirschberg, J. (2007). The Prosody of Backchannels in American English. In *Proceedings of ICPhS 16, Saarbrücken, Germany* (pp. 1065–1068).

Beňuš, Š., Trnka, M., Kuric, E., Martá, L., Gravano, A., Hirschberg, J., & Levitan, R. (2018). Prosodic entrainment and trust in human-computer interaction. In *Proceedings of the International Conference on Speech Prosody, Poznań, Poland* (pp. 220–224).

Bilous, F. R., & Krauss, R. M. (1988). Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language and Communication*, 8(3–4), 183–194.

Bock, K. (1986). Syntactic persistence in language processing. *Cognitive Psychology*, 18, 355–387.

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. retrieved 31 March 2019.

Borrie, S. A., & Delfino, C. R. (2017). Conversational Entrainment of Vocal Fry in Young Adult Female American English Speakers. *Journal of Voice*, 31(4), 1–8.

Bourhis, R., & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, ethnicity, and intergroup relations* (pp. 119–136). London: Academic Press.

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, 121(1), 41–57.

Braun, A., & Künzel. (2003). The Effect of Alcohol on Speech Prosody. In *Proceedings of ICPhS 15, Barcelona, Spain* (pp. 2645–2648).

Britain, D. (2014). Where North Meets South? Contact, Divergence and the Routinisation of the Fenland Dialect Boundary. In D. Watt & C. Llamas (Eds.), *Language, Borders and Identity* (pp. 27–43). Edinburgh: Edinburgh University Press.

Broadbent, J. M. (2008). t-to-r in West Yorkshire English. *English Language and Linguistics*, 12(1), 141–168.

Brown, G. (2014). *Y-ACCDIST: An automatic accent recognition for forensic applications*. (MA Dissertation). University of York, York.

Brown, G., & Wormald, J. (2017). Automatic sociophonetics: Exploring corpora with a forensic accent recognition system. *Journal of the Acoustical Society of America*, 142(1), 422–433.

Burland-Gibson, K. (2019). *Sociophonetic Variation and Change in a Post-Industrial , South Yorkshire Speech Community*. (Doctoral thesis). University of Sheffield, Sheffield.

Byrne, C., & Foulkes, P. (2004). The ‘mobile phone effect’ on vowel formants. *International Journal of Speech, Language and the Law*, 11(1), 83–102.

Byrne, D. (1971). *The attraction paradigm*. New York: Academic Press.

Cambier-Langeveld, T. (2014). State-of-the-art in language analysis: A response to the chapter on LADO in the Oxford handbook of language and law. *International Journal of Speech, Language and the Law*, 21(2), 371–381.

Cao, W. (2015). Phonetic Convergence of Mandarin L2 English Speakers towards Australian English. In *Proceedings of ICPhS 18, Glasgow, UK* (pp. 2–6).

Cao, W. (2018). *Short-term accommodation of Hong Kong English speakers towards native English accents and the effect of language attitudes*. (Doctoral thesis). University of York, York.

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, 40(2), 249–268.

Chirrey, D. (1999). Edinburgh: descriptive material. In P. Foulkes & G. Docherty (Eds.), *Urban voices: accent studies in the British Isles.* (pp. 223–229). London and New York: Routledge Reprint, 2014.

Clark, L., & Watson, K. (2011). Testing claims of a usage-based phonology with Liverpool English t-to-r. *English Language and Linguistics*, 15(3), 523–547.

Clark, L., & Watson, K. (2016). Phonological leveling, diffusion, and divergence: /t/ lenition in Liverpool and its hinterland. *Language Variation and Change*, 28, 31–52.

Cohen Priva, U., Edelist, L., & Gleason, E. (2017). Converging to the baseline: Corpus evidence for convergence in speech rate to interlocutor's baseline. *Journal of the Acoustical Society of America*, 141(5), 2989–2996.

Coles, F. A. (1993). /s/ Variation as Accommodation. *Kansas Working Papers in Linguistics*, 18, 31–43.

Condon, W. S., & Ogston, W. D. (1967). A segmentation of behavior. *Journal of Psychiatric Research*, 5(3), 221–235.

Coupland, N. (1984). Accommodation at work: Some phonological data and their implications. *International Journal of the Sociology of Language*, 1984(46), 49–70. Retrieved from https://doi.org/10.1515/ijsl.1984.46.49

Cowan, B. R., & Branigan, H. P. (2015). Does voice anthropomorphism affect lexical align ment in speech-based human-computer dialogue? In *Proceedings of Interspeech 2015, Dresden, Germany* (pp. 155–159).

Cowan, B. R., Branigan, H. P., Doyle, P., Edwards, J., Garaialde, D., Cabral, J., … Clark, L. (2019). What's in an accent? The impact of accented synthetic speech on lexical choice in human-machine dialogue. In *Proceedings of the 1st International Conference on Conversational User Interfaces, Dublin, Ireland* (pp. 1–8).

Cowan, B. R., Branigan, H. P., Obregón, M., Bugis, E., & Beale, R. (2015). Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human-computer dialogue. *International Journal of Human Computer Studies*, 83, 27–42.

Delvaux, V., & Soquet, A. (2007). The Influence of Ambient Speech on Adult Speech Productions through Unintentional Imitation. *Phonetica*, 64(145–173).

Devlin, T. P. (2014). *Sociophonetic Variation , Orientation and Topic in County Durham*. (Doctoral thesis). University of York, York.

Di Paolo, M., Yaeger-Dror, M., & Beckford Wassink, A. (2011). Analyzing vowels. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: a student's guide* (pp. 87–106). London: Routledge.

Dixon, J. A., Tredoux, C. G., Durrheim, K., & Foster, D. H. (1994). The role of speech accommodation and crime type in attribution of guilt. *Journal of Social Psychology*, 134(4), 465–473.

Docherty, G., & Foulkes, P. (1999). Derby and Newcastle: instrumental phonetics and variationist studies. In P. Foulkes & G. Docherty (Eds.), *Urban voices: accent studies in* *the British Isles.* (pp. 47–71). London and New York: Routledge Reprint, 2014.

Docherty, G., Foulkes, P., Milroy, J., Milroy, L., & Walshaw, D. (1997). Descriptive adequacy in phonology: A variationist perspective. *Journal of Linguistics*, 33(2), 275–310.

Douglas-Cowie, E. (1978). Linguistic code-switching in a Northern Irish village. In P. Trudgill (Ed.), *Sociolinguistic patterns in British English* (pp. 37–51). London: Edward Arnold.

Drummond, R. (2011). Glottal variation in /t/ in non-native English speech: Patterns of acquisition. *English World-Wide*, 32(3), 280–308.

Drygajlo, A., Jessen, M., Gfroerer, S., Wagner, I., Vermeulen, J., & Niemi, T. (2016). Methodological Guidelines for Best Practice in Forensic Semiautomatic and Automatic Speaker Recognition. Frankfurt: Verlag für Polizeiwissenschaft.

Earnshaw, K. (forthcoming). Examining the implications of speech accommodation for forensic speaker comparison casework : A case study of the West Yorkshire FACE vowel. *Journal of Phonetics*.

Earnshaw, K. (2013). *The relationship between frequency of use and perceptual awareness of TH-fronting*. (BA Dissertation). University of Sheffield, Sheffield.

Earnshaw, K. (2014). *Assessing the Discriminatory Power of /t/ and /k/ for Forensic Speaker Comparison using a Likelihood Ratio Approach*. (MSc Dissertation). University of York, York.

Earnshaw, K., & Gold, E. (2018). Variation in the FACE vowel across West Yorkshire: Implications for forensic speaker comparisons. In *Proceedings of Interspeech 2018, Hyderabad, India* (pp. 2743–2747).

Earnshaw, K., & Gold, E. (2019). “We don’t pronounce our T’s around here": Realisations of /t/ in West Yorkshire English. In *Proceedings of ICPhS 19, Melbourne, Australia* (pp. 2655–2659).

Easther, A. (1883). *A glossary of the dialect of Almondbury and Huddersfield*. Vaduz: Kraus Reprint, 1965.

Edlund, J., Heldner, M., & Hirschberg, J. (2009). Pause and gap length in face-to-face interaction. In *Proceedings of Interspeech 2009, Brighton, UK* (pp. 2779–2782).

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America*, 121(6), 3814.

Fabricius, A. (2002). Ongoing change in modern RP: Evidence for the disappearing stigma of t-glottalling. *English World-Wide*, 23(1), 115–136.

Finlayson, I., Lickley, R., & Corley, M. (2012). Convergence of speech rate: Interactive alignment beyond representation’. In *Proceedings of the 25th Annual CUNY Conference on Human Sentence Processing, New York, USA* (Vol. 11, p. 24).

First Bus North. (2020). Tech the Bus - Smarter Travel | Leeds |. Retrieved 13 March 2020, from https://www.firstgroup.com/leeds/plan-journey/tech-bus-smarter-travel

Fitzpatrick, M., Kim, J., & Davis, C. (2015). The effect of seeing the interlocutor on auditory and visual speech production in noise. *Speech Communication*, 74, 37–51.

Flanagan, J. L. (1955). A difference limen for vowel formant frequency. *Journal of the Acoustical Society of America*, 27(3), 613–617.

Flynn, N. (2011). Comparing vowel formant normalisation procedures. *York Papers in Linguistics Series 2*, (11).

Forensic Science Regulator. (2021). *Codes of Practice and Conduct*.

Foulkes, P., & French, J. P. (2012). Forensic Speaker Comparison: A Linguistic–Acoustic Perspective. In L. M. Solan & P. M. Tiersma (Eds.), *Oxford Handbook of Language and Law* (pp. 557–572). Oxford: Oxford University Press.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct–realist perspective. *Journal of Phonetics*, 14(1), 3–28.

French, J. P. (2017). A Developmental History of Forensic Speaker Comparison in the UK. *English Phonetics*, 271–286.

Fuscone, S., Favre, B., & Prévot, L. (2018). Replicating Speech Rate Convergence Experiments on the Switchboard Corpus. In *Proceedings of LREC’18, Miyazaki, Japan.*

Gambi, C., & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology*, 4, 1–9.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition*, 53(3), 181–215.

Gauder, L., Reartes, M., Gálvez, R. H., Běnuš, Š., & Gravano, A. (2018). Testing the effects of acoustic/prosodic entrainment on user behavior at the dialog-act level. In *Proceedings of the International Conference on Speech Prosody, Poznań, Poland* (pp. 374–378).

Gavaldà, N. (2016). Individual Variation in Allophonic Processes of /T/ in Standard Southern British English. *International Journal of Speech, Language and the Law*, 23(1), 43–69.

Gessinger, I., Mobius, B., Fakhar, N., Raveh, E., & Steiner, I. (2019). A Wizard-of-Oz experiment to study phonetic accommodation in human-computer interaction. In *Proceedings of*

ICPhS 19, Melbourne, Australia. (pp. 1475–1479).

Gijssels, T., Casasanto, L. S., Jasmin, K., Hagoort, P., & Casasanto, D. (2016). Speech Accommodation Without Priming: The Case of Pitch. *Discourse Processes*, 53(4), 233–251.

Giles, H. (1973). Accent Mobility: A Model and Some Data. *Anthropological Linguistics*, 15(2), 87–105.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of Accommodation: Developments in Applied Sociolinguistics* (pp. 1–68). Cambridge: Cambridge University Press.

Giles, H., Mulac, A., Bradac, J. J., & Johnson, P. (1987). Speech Accommodation Theory: The First Decade and Beyond. *Annals of the International Communication Association*, 10(1), 13–48.

Giles, H., & Powesland, P. F. (1975). *Speech style and social evaluation*. London and New York: Academic Press.

Giles, H., Taylor, D. M., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society*, 177–192.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), San Francisco, CA, USA* (Vol. 1, pp. 517–520).

Gold, E. (2014). *Calculating likelihood ratios for forensic speaker comparisons using phonetic and linguistic parameters*. (Doctoral thesis). University of York, York.

Gold, E., & French, J. P. (2011). International practices in forensic speaker comparisons. *International Journal of Speech, Language and the Law*, 18(2), 293–307.

Gold, E., & French, J. P. (2019). International practices in forensic speaker comparisons: Second survey. *International Journal of Speech, Language and the Law*, 26(1), 1–20.

Gold, E., French, J. P., & Harrison, P. (2013a). Clicking behavior as a possible speaker discriminant in English. *Journal of the International Phonetic Association*, 43(3), 339–349.

Gold, E., French, J. P., & Harrison, P. (2013b). Examining long-term formant distributions as a discriminant in forensic speaker comparisons under a likelihood ratio framework. In *Proceedings of Meetings on Acoustics, Montreal, Canada* (Vol. 19, pp. 1–8).

Gold, E., Kirchhübel, C., Earnshaw, K., & Ross, S. (forthcoming). Regional Variation in British English Voice Quality. *English World-Wide*.

Gold, E., Ross, S., & Earnshaw, K. (2018). The 'West Yorkshire Regional English Database': Investigations into the generalizability of reference populations for forensic speaker comparison casework. In *Proceedings of Interspeech 2018, Hyderabad, India* (pp. 2748–2752).

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279.

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin and Review*, 11(4), 716–722.

Gráczi, T. E., & Kohári, A. (2014). Multiple bursts in Hungarian voiceless plosives and VOT measurements. In *Proceedings of the 10th International Seminar on Speech Production (ISSP), Cologne, Germany.* (pp. 158–161).

Greenwald, A. G., Mcghee, D. E., & Schwartz, J. L. K. (1998). Pleasant Vs Unpleasant Associations. *Journal of Personality and Social Psychology*, 74(6), 1464–1480.

Gregory, S. W. (1990). Analysis of fundamental frequency reveals covariation in interview partners' speech. *Journal of Nonverbal Behavior*, 14(4), 237–251.

Grice, M., & Barry, W. (1991). Problems of transcription and labelling in the specification of segmental and prosodic structure. In *Proceedings of ICPhS 12, Aix-en-Provence, France* (pp. 66–69).

Haddican, B., Foulkes, P., Hughes, V., & Richards, H. (2013). Interaction of social and linguistic constraints on two vowel changes in northern England. *Language Variation and Change*, 25(3), 371–403.

Hale, J. L., & Burgoon, J. K. (1984). Models of reactions to changes in nonverbal immediacy. *Journal of Nonverbal Behavior*, 8(4), 287–314.

Harris, J., & Kaye, J. (1990). A tale of two cities: London glottalling and New York City tapping. *The Linguistic Review*, 7, 251–274.

Harrison, P. (2019a). f0Analysis.praat Version 1.9.4. [Praat script].

Harrison, P. (2019b). formantLog.praat Version 1.7.4. [Praat script].

Hazan, V., Tuomainen, O., Kim, J., Davis, C., Sciences, P., & Sydney, W. (2019). The effect of visual cues on speech characteristics of older and younger adults in an interactive task. In *Proceedings of ICPhS 19, Melbourne, Australia* (pp. 815–819).

Heldner, M., Edlund, J., Hirschberg, J., & York, N. (2010). Pitch similarity in the vicinity of backchannels. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association, Makuhari, Japan.* (pp. 3054–3057). Makuhari, Japan.

Hickey, R. (1999). Dublin English: current changes and their motivations. In P. Foulkes & G. Docherty (Eds.), *Urban voices: accent studies in the British Isles.* (pp. 265–281). London and New York: Routledge Reprint, 2014.

Hughes, A., Trudgill, P., & Watt, D. (2012). *English accents & dialects : an introduction to social and regional varieties of English in the British Isles* (5th ed.). London: Routledge.

Hughes, V., & Foulkes, P. (2017). What is the relevant population? Considerations for the computation of likelihood ratios in forensic voice comparison. In *Proceedings of Interspeech 2017, Stockolm, Sweden* (pp. 3772–3776).

Hughes, V., McDougall, K., & Foulkes, P. (2009). Diphthong dynamics in unscripted speech. *Presented at IAFPA, Cambridge, UK.*

Hughes, V., Wood, S., & Foulkes, P. (2016). Strength of forensic voice comparison evidence from the acoustics of filled pauses. *International Journal of Speech, Language and the Law*, 23(1), 99–132.

Ibrahim, O., Skantze, G., Stoll, S., & Dellwo, V. (2019). Fundamental frequency accommodation in multi-party human-robot game interactions: The effect of winning or losing. In *Proceedings of Interspeech 2019, Graz, Austria* (pp. 3980–3984).

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, 22(1), 39–44.

Jacewicz, E., Fox, R., & Salmons, J. (2011). Vowel change across three age groups of speakers in three regional varieties of American English. *Journal of Phonetics*, 39(4), 683–693. Retrieved from https://doi.org/10.1016/j.wocn.2011.07.003.Vowel

Jones, W. E. (1952). The Definite Article in Living Yorkshire Dialect. *Leeds Studies in English*, 7–8, 81–91.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, 2(1), 125–156.

Kirchhübel, C., Howard, D. M., & Stedmon, A. W. (2011). Acoustic correlates of speech when under stress: Research, methods and future directions. *International Journal of Speech,* *Language and the Law*, 18(1), 75–98.

Kirkham, S., & Moore, E. (2016). Constructing social meaning in political discourse: Phonetic variation and verb processes in Ed Miliband's speeches. *Language in Society*, 45(1), 87–111.

Künzel, H. J. (2001). Beware of the 'telephone effect': The influence of telephone transmission on the measurement of formant frequencies. *International Journal of Speech, Language and the Law*, 8(1), 80–99.

Kurtić, E., & Gorisch, J. (2018). F0 accommodation and turn competition in overlapping talk. *Journal of Phonetics*, 71, 376–394.

Kurtić, E., Wells, B., Brown, G. J., Kempton, T., & Aker, A. (2012). A corpus of spontaneous multi-party conversation in Bosnian Serbo-Croatian and British English. In *Proceedings of LREC'12, Istanbul, Turkey* (Vol. 2, pp. 1323–1327).

Labov, W. (1966). *The social stratification of English in New York City*. Washington DC: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W., Ash, S., Baranowski, M., Nagy, N., & Ravindranath, M. (2006). Listeners' sensitivity to the frequency of sociolinguistic variables. *University of Pennsylvania Working Papers in Linguistics: Selected Papers from NWAV 34*, 12(2), 105–129.

Labov, W., Ash, S., Ravindranath, M., Weldon, T., Baranowski, M., & Nagy, N. (2011). Properties of the sociolinguistic monitor. *Journal of Sociolinguistics*, 15(4), 431–463.

Ladefoged, P. (2001). *A Course in Phonetics* (4th ed.). Fort Worth: Harcourt.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Leach, H. (2018). *Sociophonetic variation in Stoke-on-Trent's pottery industry*. (Doctoral thesis). University of Sheffield, Sheffield.

Lee, C. C., Black, M., Katsamanis, A., Lammert, A., Baucom, B., Christensen, A., … Narayanan, S. (2010). Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples. In *Proceedings of Interspeech 2010, Makuhari, Japan* (pp. 793–796).

Leeds City Council. (2020). Leeds economy. Retrieved 12 October 2020, from https://www.leeds.gov.uk/business/leeds-economy

Leemann, A., Blaxter, T., Britain, D., & Earnshaw, K. (2019). The FACE of Change in English Dialects: 1950 v. 2018. In *Proceedings of ICPhS 19, Melbourne, Australia.* (pp. 373–377).

Leemann, A., Kolly, M. J., & Britain, D. (2018). The English Dialects App: The creation of a crowdsourced dialect corpus. *Ampersand*, 5, 1–17.

Lennes, M. (2017). SpeCT - Speech Corpus Toolkit for Praat (v1.0.0). First release on GitHub.

Levitan, R., Beňuš, Š., Gálvez, R. H., Gravano, A., Savoretti, F., Trnka, M., … Hirschberg, J. (2016). Implementing acoustic-prosodic entrainment in a conversational avatar. In *Proceedings of Interspeech 2016, San Francisco, CA, USA* (pp. 1166–1170).

Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association, Florence, Italy* (pp. 3081–3084).

Levon, E., & Fox, S. (2014). *Social Salience and the Sociolinguistic Monitor: A Case Study of ING and TH-fronting in Britain*. *Journal of English Linguistics* (Vol. 42).

Lewandowski, N., & Jilka, M. (2019). Phonetic Convergence, Language Talent, Personality and Attention. *Frontiers in Communication*, 4, 1–19.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revisited. *Cognition*, 21, 1–36.

Llamas, C. (2007). ‘A place between places’: Language and identities in a border town. *Language in Society*, 36(4), 579–604.

Llamas, C., Watt, D., & Johnson, D. E. (2009). Linguistic accommodation and the salience of national identity markers in a border town. *Journal of Language and Social Psychology*, 28(4), 381–407.

Love, J., & Walker, A. (2013). Football versus football: Effect of topic on /r/ realization in American and English sports fans. *Language and Speech*, 56(4), 443–460.

Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior*, 34(6), 419–426.

Mathisen, A. G. (1999). Sandwell, West Midlands: ambiguous perspectives on gender patterns and models of change. In P. Foulkes & G. Docherty (Eds.), *Urban voices: accent studies in the British Isles.* (pp. 107–123). London and New York: Routledge Reprint, 2014.

Maurer, R. E., & Tindall, J. H. (1983). Effect of postural congruence on client’s perception of counselor empathy. *Journal of Counseling Psychology*, 30(2), 158–163.

McCafferty, K. (1999). (London) Derry: between Ulster and local speech - class, ethnicity and language change. In P. Foulkes & G. Docherty (Eds.), *Urban voices: accent studies in the British Isles.* (pp. 246–264). London and New York: Routledge Reprint, 2014.

McDougall, K. (2004). Speaker-specific formant dynamics: An experiment on Australian English /aI/. *International Journal of Speech, Language and the Law*, 11(1), 103–130.

Mees, I. M., & Collins, B. (1999). Cardiff: a real-time study of glottalization. In P. Foulkes & G. Docherty (Eds.), *Urban voices: accent studies in the British Isles.* (pp. 185–202). London and New York: Routledge Reprint, 2014.

Meyerhoff, M. (2011). *Introducing Sociolinguistics* (2nd ed.). USA and Canada: Routledge.

Milroy, J., Milroy, L., & Hartley, S. (1994). Local and Supra-Local Change in British English - The Case of Glottalisation. *English World-Wide*, 15(1), 1–33.

Morrison, G. S., Sahito, F. H., Jardine, G., Djokic, D., Clavet, S., Berghs, S., & Goemans Dorny, C. (2016). INTERPOL survey of the use of speaker identification by law enforcement agencies. *Forensic Science International*, 263, 92–100.

Mukherjee, S., Ausilio, A. D., Nguyen, N., Fadiga, L., Badino, L., & Italiano, I. (2017). The Relationship between F0 Synchrony and Speech Convergence in Dyadic Interaction. In *Proceedings of Interspeech 2017, Stockolm, Sweden* (pp. 2341–2345).

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, 21(4), 422–432.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, 32(5), 790–804.

Nenkova, A., Gravano, A., & Hirschberg, J. (2008). High frequency word entrainment in spoken dialogue. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. ACL HLT 2008. Columbus, Ohio, USA.* (pp. 169–172).

Newbrook, M. (1999). West Wirral: norms, self-reports and usage. In P. Foulkes & G. Docherty (Eds.), *Urban voices: accent studies in the British Isles.* (pp. 90–106). London and New York: Routledge Reprint, 2014.

Nguyen, N., & Delvaux, V. (2015). Role of imitation in the emergence of phonological systems. *Journal of Phonetics*, 53, 46–54.

Nielsen, K. (2007). Implicit Phonetic Imitation Is Constrained By Phonemic Contrast. In *Proceedings of ICPhS 16, Saarbrücken, Germany* (pp. 1961–1964).

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2), 132–142.

Nolan, F. (1991). Forensic phonetics. *Journal of Linguistics*, 27(2), 483–493.

Nolan, F. (2012). Degrees of freedom in speech production: An argument for native speakers in LADO. *International Journal of Speech, Language and the Law*, 19(2), 263–289.

Nolan, F., McDougall, K., de Jong, G., & Hudson, T. (2009). The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech, Language and the Law*, 16(1), 31–57.

Office for National Statistics. (2012a). 2011 Census: Ethnic group, local authorities in England and Wales. Retrieved 12 October 2020, from https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/2011censuskeystatisticsforlocalauthoritiesinenglandandwales

Office for National Statistics. (2012b). 2011 Census: Industry, local authorities in England and Wales. Retrieved 12 October 2020, from https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/2011censuskeystatisticsforlocalauthoritiesinenglandandwales

Office for National Statistics. (2013). 2011 Census: Main language (detailed), local authorities in England and Wales. Retrieved 12 October 2020, from https://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-and-quick-statistics-for-wards-and-output-areas-in-england-and-wales/rft-qs204ew.xls

Office for National Statistics. (2020). Population Estimates for UK, England and Wales, Scotland and Northern Ireland. Retrieved 20 October 2020, from https://www.ons.gov.uk/datasets/mid-year-pop-est/editions/mid-2019-april-2020-geography/versions/1

Orton, H., & Dieth, E. (1962). *Survey of English Dialects*. Leeds: J. Arnold & Sons.

Papp, V. (2009). The effects of heroin on the spectral quality of vowels. *Presented at IAFPA, Cambridge, UK*.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the*

Acoustical Society of America, 119(4), 2382–93.

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, 40(1), 190–197.

Pardo, J. S., Jay, I. C., Hoshino, R., Hasbun, S. M., Sowemimo-Coker, C., & Krauss, R. M. (2013a). Influence of Role-Switching on Phonetic Convergence in Conversation. *Discourse Processes*, 50(4), 276–300.

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception & Psychophysics*, 72(8), 2254–2264.

Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013b). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language*, 69(3), 183–195.

Pardo, J. S., Urmanche, A., Gash, H., Wiener, J., Mason, N., Wilman, S., … Decker, A. (2018a). The Montclair Map Task Corpus of Conversations in English. Linguistic data corpus available at http://digitalcommons. montclair.edu/mmt_corpus.

Pardo, J. S., Urmanche, A., Gash, H., Wiener, J., Mason, N., Wilman, S., … Decker, A. (2019). The Montclair Map Task: Balance, Efficacy, and Efficiency in Conversational Interaction. *Language and Speech*, 62(2), 378–398.

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, and Psychophysics*, 79(2), 637–659.

Pardo, J. S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., & Ward, M. (2018b). A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics*, 69, 1–11.

Petyt, K. M. (1985). *Dialect and accent in industrial West Yorkshire*. Amsterdam: John Benjamins Publishing Company.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169–226.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347.

Purnell, T. C. (2009). Convergence and contact in Milwaukee: Evidence from select African American and White vowel space features. *Journal of Language and Social Psychology*, 28(4), 408–427.

R Core Team. (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rickford, J., & McNair-Knox, F. (1994). Addressee- and topic- influenced style shift: A quantitative sociolinguistic study. In D. Biber & E. Finegan (Eds.), *Perspectives on Register: Situating Register Variation within Sociolinguistics* (pp. 235–276). Oxford: Oxford University Press.

Rogers, H. (2000). *The Sounds of Language: An Introduction to Phonetics*. London and New York: Routledge Reprint, 2013.

Rose, P. (2003). The technical comparison of forensic voice samples. In H. Selby & I. Freckelton (Eds.), *Expert Evidence 99*. Sydney: Thomas Lawbook Co.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25(4), 421–436.

Schleef, E. (2013). Glottal replacement of /t/ in two British capitals: Effects of word frequency and morphological compositionality. *Language Variation and Change*, 25(2), 201–223.

Schweitzer, A., & Lewandowski, N. (2014). Social factors in convergence of F1 and F2 in spontaneous speech. In *Proceedings of ISSP 2014, Cologne, Germany* (pp. 391–394).

Sharma, D. (2011). Style repertoire and social change in British Asian English. *Journal of Sociolinguistics*, 15(4), 464–492.

Shepard, C. A., Giles, H., & LePoire, B. A. (2001). Communication accommodation theory. In W. P. Robinson & H. Giles (Eds.), *The New Handbook of Language and Social Psychology* (pp. 33–56). New York: Wiley.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, 66(3), 422–429.

Smith, J., & Holmes-Elliott, S. (2015). Parenting style: From preschool to preadolescence in the acquisition of variation. *Presented at NWAV 44, University of Toronto, Canada*.

Smith, J., & Holmes-Elliott, S. (2017). The unstoppable glottal: tracking rapid change in an iconic British variable. *English Language and Linguistics*, 22(3), 323–355.

Solanki, V. J. (2017). *Brains in dialogue: Investigating accommodation in live conversational speech for both speech and EEG data.* (Doctoral thesis). University of Glasgow, Glasgow.

Solanki, V. J., Vinciarelli, A., Stuart-Smith, J., & Smith, R. H. (2015). Utilising Hidden Markov Modelling for the Assessment of Accommodation in Conversational Speech. In *Proceedings of ICPhS 2015, Glasgow, UK* (pp. 10–14).

Sóskuthy, M. (2017). *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction*.

Stamp, R., Schembri, A., Evans, B. G., & Cormier, K. (2015). Regional Sign Language Varieties in Contact: Investigating Patterns of Accommodation. *Journal of Deaf Studies and Deaf Education*, 1–13.

Staum Casasanto, L., Jasmin, K., & Casasanto, D. (2010). Virtually accommodating: Speech rate accommodation to a virtual interlocutor. In *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society, Portland, Oregon, USA.* (Vol. 1, pp. 127–132).

Stevens, L., & French, J. P. (2012). Voice quality in studio quality and telephone transmitted recordings. *Presented at the British Association of Academic Phonetics Conference, Leeds, UK.*

Stoddart, J., Upton, C., & Widdowson, J. D. A. (1999). Sheffield dialect in the 1990s: revisiting the concept of NORMs. In P. Foulkes & G. Docherty (Eds.), *Urban voices: accent studies in the British Isles.* (pp. 72–89). London and New York: Routledge Reprint, 2014.

Strycharczuk, P., López-ibáñez, M., Brown, G., & Leemann, A. (2020). General Northern English . Exploring Regional Variation in the North of England With Machine Learning. *Frontiers in Artificial Intelligence*, 3, 1–18.

Stuart-Smith, J. (1999a). Glasgow: accent and voice quality. In P. Foulkes & G. Docherty (Eds.), *Urban voices: accent studies in the British Isles.* (pp. 203–222). London and New York: Routledge Reprint, 2014.

Stuart-Smith, J. (1999b). Glottals Past and Present: A Study of T-glottaling in Glaswegian. *Leeds Studies in English*, 181–204.

Stuart-Smith, J., Timmins, C., & Alam, F. (2011). Hybridity and ethnic accents. A sociophonetic analysis of 'Glaswasian'. In F. Gregersen, J. Parrot, & P. Quist (Eds.), *Language Variation - European Perspectives III: Selected papers from the 5th International Conference on Language Variation in Europe (ICLaVE 5)* (pp. 43–58). Amsterdam: John Benjamins Publishing Company.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 79(4), 1086–1100.

Tamminga, M. (2016). Individual differences in convergence and sound change. *Presented at NWAV 45, University of Pennsylvania, PA*.

Taylor, P. J., & Thomas, S. (2008). Linguistic Style Matching and Negotiation Outcome. *Negotiation and Conflict Management Research*, 1(3), 263–281.

The International Association for Forensic Phonetics and Acoustics. (2018). Research grant applications welcome - 2018/19. Retrieved 6 January 2019, from https://www.iafpa.net/the-association/news/

Thomas, E. R., Kendall, T., Yaeger-Dror, M., & Kretzschmar, W. (2007). Two things sociolinguists should know: Software packages for vowel normalization, and accessing linguistic atlas data. *Workshop Presented at NWAV 36, University of Pennsylvania, PA*.

Tollfree, L. (1999). South East London English: discrete versus continuous modelling of consonantal reduction. In P. Foulkes & G. Docherty (Eds.), *Urban voices: accent studies in the British Isles.* (pp. 163–184). London and New York: Routledge Reprint, 2014.

Traunmüller, H. (1997). Auditory scales of frequency representation. Retrieved 1 November 2018, from http://www.ling.su.se/staff/hartmut/bark.htm

Trudgill, P. (1974). *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Trudgill, P. (1981). Linguistic accommodation: Sociolinguistic observations on a sociopsychological theory. In R. Hendrick, C. Masek, & M. F. Miller (Eds.), *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistics Society.

Trudgill, P. (1986). *Dialects in contact*. Oxford: Blackwell.

Trudgill, P. (1999). Norwich: endogenous and exogenous linguistic change. In P. Foulkes & G. Docherty (Eds.), *Urban voices: accent studies in the British Isles.* (pp. 124–140). London and New York: Routledge Reprint, 2014.

Turton, D. (2017). t-flapping in present-day Northern English : implications for sociophonological change. *Presented at the Third Symposium on Historical Phonology, University of Edinburgh, UK*.

van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The wildcat corpus of native-and foreign-accented english: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, 53(4), 510–540.

Walker, A. (2014). *Crossing Oceans with Voices and Ears: Second Dialect Acquisition and Topic-Based Shifting in Production and Perception*. (Doctoral thesis). Ohio State University, Ohio.

Walker, A. (2019). The Role of Dialect Experience in Topic-Based Shifts in Speech Production. *Language Variation and Change*, 31(2), 135–163.

Walker, A., & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology*, 6, 1–18.

Watson, K. (2006). Phonological resistance and innovation in the North-West of England. *English Today*, 22(2), 55–61.

Watt, D., Llamas, C., French, J. P., Braun, A., & Robertson, D. (2018). A new corpus of Northern Englishes: Building the TUULS database for sociolinguistic and forensic research on phonetic variation in the Northeast of England. *Presented at The 8th Northern Englishes Workshop, Newcastle, UK*.

Watt, D., & Milroy, L. (1999). Patterns of variation and change in three Newcastle vowels: is this dialect levelling? In P. Foulkes & G. Docherty (Eds.), *Urban voices: accent studies in the British Isles.* (pp. 25–46). London and New York: Routledge Reprint, 2014.

Watt, D., & Tillotson, J. (2001). A spectrographic analysis of vowel fronting in Bradford English. *English World-Wide*, 22(2), 269–302. Retrieved from https://doi.org/10.1039/RG9446800197

Weatherholtz, K., Campbell-Kibler, K., & Jaeger, T. F. (2014). Socially-mediated syntactic alignment. *Language Variation and Change*, 26(3), 387–420.

Wells, J. C. (1982). *Accents of English.* (1st ed., Vol. 1). Cambridge: Cambridge University Press.

Wells, J. C. (1985). English accents in England. In P. Trudgill (Ed.), *Language in the British Isles* (pp. 55–69). Cambridge: Cambridge University Press.

Whisker-Taylor, K., & Clark, L. (2019). Yorkshire Assimilation: Exploring the Production and Perception of a Geographically Restricted Variable. *Journal of English Linguistics*, 47(3), 221–248.

Wilhelm, S. (2018). Segmental and suprasegmental change in North West Yorkshire – a new case of supralocalisation? *Corela*, (HS-24).

Williams, A., & Kerswill, P. (1999). Dialect levelling: change and continuity in Milton Keynes, Reading and Hull. In P. Foulkes & G. Docherty (Eds.), *Urban voices: accent studies in the British Isles.* (pp. 141–162). London and New York: Routledge Reprint, 2014.

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications. [http://arxiv.org/pdf/1308.5499.pdf]*.

Wolfson, N. (1976). Speech Events and Natural Speech. *Language in Society*, (5), 189–209.

Wood, S. (2006). *Generalized additive models: an introduction with R*. Boca Raton: CRC Press.

Wormald, J. (2016). *Regional Variation in Panjabi-English*. (Doctoral thesis). University of York, York.

Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic Imitation from an Individual-Difference Perspective: Subjective Attitude, Personality and 'Autistic' Traits. *PLoS ONE*, 8(9).

Zellers, M., & Schweitzer, A. (2017). An investigation of pitch matching across adjacent turns in a corpus of spontaneous German. In *Proceedings of Interspeech 2017, Stockolm, Sweden* (pp. 2336–2340).