Speaker Identification Using Laughter in a Close Social Network

Forensically-relevant research on laughter is extremely limited in the literature, however, experts have reported analysing laughter in forensic speaker comparison casework (Gold and French 2011). This paper describes a preliminary investigation into the potential speaker-specificity of laughter. A close social network of 7 undergraduate university students took part in an open speaker identification task containing 4-second samples of their laughter. Overall, the network members performed much worse than in a similar study using speech samples (see Foulkes and Barron 2000), as each network member identified only one speaker correctly. The largest number of correct identifications of any speaker was three, while another three of the network members were never correctly identified. Previous studies that have also investigated laughter using voice line-ups have reported higher identification rates (Philippon et al. 2013; Yarmey 2004). The differences between the results of the present study and previous studies may be explained by qualitative and quantitative differences in the laughter samples used, particularly differences in voicing and sample length. This suggests that longer samples of specifically voiced laughter may facilitate higher naïve speaker identification rates. Further research is still needed on the possible speaker-specificity of voiced laughter but it may have the potential to be developed for use as a speaker discriminant in forensic phonetic casework.

KEYWORDS: LAUGHTER, NAÏVE SPEAKER IDENTIFICATION, SPEAKER-SPECIFICITY.

1.0 INTRODUCTION

This paper considers the speaker-specificity of laughter. Whilst some scholars have suggested that laughter may influence people's memory of others (see Armony, Chochol, Fecteau and Belin 2007), there has been little forensically-relevant research examining how listeners perform when carrying out naïve speaker identification using laughter (cf. Yarmey 2004; Philippon, Randall and Cherryman 2013). Such research would benefit the field of forensic phonetics, as it would indicate whether the use of voice line-ups could be extended to cases involving ear-witnesses hearing laughter rather than speech. Additionally, because the ability to identify speakers based on samples suggests that something in those samples is speaker-specific (Foulkes and Barron 2000), such research would also indicate whether laughter would be worthy of further research with an aim to develop it as a non-linguistic speaker discriminant for use in forensic speaker comparison casework. This would be useful because a small number of forensic phoneticians have reported analysing laughter in forensic speaker comparison casework (Gold and French 2011), seemingly without the necessary research to support their analysis. Similarly, Hirson (1995) explicitly notes how, in a specific case involving only laughter in the disputed sample and only speech in the reference samples, the lack of research into the speaker-specificity of laughter itself and the potential correspondence between acoustic parameters common to both laughter and speech limited the strength of the conclusions that could be made about speaker identity. The present study therefore aims to make one of the first preliminary investigations into the speaker-specificity of laughter by investigating whether members of a close social network can identify one another from samples of laughter in a speaker identification task.

2.0 LAUGHTER

Laughter can be described in terms of both its audio and visual properties (see Cosentino, Sessa and Takanishi 2016: 150). However, the present study is concerned only with laughter's audio properties (cf. Petridis, Martinez and Pantic 2013), which might often be found written orthographically as 'ha' or 'haha'. In addition, whilst different scholars have used contradictory terminology to describe laughter's auditory properties in differing levels of detail, for the purposes of this study, laughter is understood to consist of calls, bouts, and episodes (Cosentino et al. 2016; cf. for instance, Mowrer, LaPointe and Case 1987) which can be unvoiced, voiced, and mixed (Bachorowski, Smoski and Owren 2001; see Figures 2, 3 and 4 below). A single laughter *call* can be thought of in terms of the orthographic ‘ha’, though this oversimplifies the phonetic details surrounding voicing. Voiced calls are composed of aspiration followed by a vocalic element, whereas voiceless calls consist mostly just of aspiration. Bachorowski et al. (2001) suggest this vocalic element typically resembles a schwa vowel, whilst Szameitat, Darwin, Szameitat, Wildgruber and Alter (2011) report higher acoustic values in the first formant frequency that would suggest a more articulate vocalisation closely resembling /a/. At least one such call is needed to form a bout of laughter (e.g. ‘haha’) and at least one bout of laughter is needed to form an episode of laughter (e.g. ‘ha haha’). However, there are no objective criteria for deciding where one bout ends and another begins or for deciding if two or more bouts constitute one episode or several.

Laughter has received much attention from those working in conversation analysis (hereafter, CA). A fundamental aim in CA is to discover how conversational interactions are structured (see Sidnell 2010). Thus, CA research on laughter has sought to explain how laughter might structure interaction. The literature is extensive, ranging from book-length accounts (Glenn and Holt 2013) to articles covering, for instance, how laughter can initiate playful talk (Holt 2016), how laughter occurs in similar positions in the sign language of deaf individuals and the speech of normally-hearing individuals (Provine and Emmorey 2006), and how antiphonal laughter (i.e. laughter overlapping with an interlocutor’s speech) increases with the development of friendships (Smoski and Bachorowski 2003a; 2003b). As such research focuses on laughter’s conversational functions rather than the actual phonetic properties of laughter itself, it is not immediately relevant to the issue of speaker-specificity or forensic phonetics generally. However, some studies that employ CA ideas but move away from the traditional qualitative approach of CA (see Drew 2008:136--137) to include quantitative and acoustic analyses are slightly more relevant. For instance, Vettin and Todt (2004) suggest that individuals produce fewer laughter calls in laughter that follows their own conversational turns than laughter that follows their interlocutor's turns. Findings such as these, could eventually contribute to addressing the intra-speaker variability of laughter.

Some of the above studies have analysed how the acoustic properties of laughter differ based on its function. More relevant to the speaker-specificity of laughter, though, are some preliminary indications of how various speaker demographics correlate with acoustic properties of laughter. For instance, autistic children appear to produce more voiced laughter than non-autistic children (Hudenko, Stone and Bachorowski. 2009), whilst there appears to be little difference between congenitally deaf and normally-hearing college students' laughter (Makagon, Funayama, and Owren 2008). Bachorowski et al. (2001) report that a discriminant function analysis correctly classified a sample of 97 subjects (45 males, 52 females) by sex 72.6% of the time using the fundamental frequency and the first three formant frequencies in voiced laughter. Bachorowski et al. (2001) also briefly note, but provide few further details, that the same parameters were less than 50% accurate in classifying individual speakers. Again, whilst making useful contributions to our understanding of the relationship between laughter and identity in terms of broad speaker demographics, such research does not address the question of whether laughter is specific to individuals.

Only two previous studies by Yarmey (2004) and Philippon et al. (2013) have in fact begun to address the speaker-specificity of laughter, and several problematic aspects of both studies' methodologies justify the existence of the present study. Philippon et al. (2013) conducted an open-set experimental voice line-up in both target-absent and target-present conditions. The line-up consisted of 15-second samples of voiced laughter from 6 foil speakers and 1 target speaker. These samples were however derived from the speakers acting out an interaction from a script which included several occasions where speakers were directed to laugh. Furthermore, the use of purely voiced laughter neglects the possibility of investigating whether voiceless laughter may be speaker-specific. 15 out of 32 listeners correctly judged the line-up, as 4 out of 16 listeners presented with the target-present condition correctly identified the target speaker and 11 out of 16 listeners presented with the target-absent condition correctly rejected the line-up. Whilst this equates to a 47% identification rate, which Philippon et al. (2013) consider to be suggestive of voiced laughter being speaker-specific, it is notable that this identification rate is mostly a result of the listeners correctly rejecting the voice line-up when the target was absent. Calculating the identification rate on the basis of correct identifications of the target speaker results in a lesser rate of 25%. Of course, being able to successfully reject foil speakers does serve as some evidence for the speaker-specificity of voiced laughter. However, as the listeners seemed to be better at rejecting the target-absent line-ups than identifying the target speaker when present, perhaps voiced laughter has speaker-specific properties to the extent that naïve listeners can use it more effectively for elimination than identification. Yarmey (2004) conducted a speaker identification task in which 43 listeners were required to listen to 1-second samples of laughter from 8 speakers (4 familiar speakers and 4 unfamiliar speakers). Similar to Philippon et al. (2013), the samples consisted of unnatural acted laughter produced on cue to the prescribed length of 1 second. Listeners had to state whether they recognised the speakers and, if they did recognise them, try to identify them by name. 112 out of 172 (65%) of listeners' answers to familiar speaker samples correctly indicated that they recognised them. 123 out of 172 (72%) of listeners' answers to unfamiliar speaker samples correctly indicated that they did not recognise them. Listeners were slightly less successful in identifying by name the familiar speakers they recognised, providing 89 out of 172 (52%) correct identifications, which echoes the abovementioned possibility of laughter being more effective in elimination than identification. Given the shorter 1 second samples of laughter, it is interesting that the identification rate of 52% is more than double the 25% rate reported in Philippon et al. (2013). However, as Yarmey (2004) does not provide any information about the voicing of the laughter samples, it is difficult to interpret the significance of this result in relation to the suggestion by Philippon et al. (2013) that voiced laughter may facilitate identification. Besides, Yarmey (2004) presented the listeners with a mostly closed-set task, because they were asked to confirm their familiarity and unfamiliarity with the 8 speakers before they took part. The possibility of the listeners using that information to deduce the speakers' identities, rather than relying solely on the information in the laughter samples, may therefore account for the higher identification rates. Irrespective of methodologies, the identification rates reported for both Yarmey (2004) and Philippon et al. (2013) are lower than those reported in similar studies of speech. For instance, Foulkes and Barron (2000) report that listeners were able to identify familiar speakers, who were members of the listeners' close social network, at a rate of 68%, which suggests that naïve speaker identification is more successful using samples of speech rather than laughter. Further evidence relating to this matter can be found in the present study, which aims to improve on the shortcomings of Philippon et al. (2013) and Yarmey (2004) by using elicited, and therefore potentially more natural laughter, whilst also focusing not just on voiced laughter but also voiceless and mixed laughter, and by presenting listeners with a relatively more open-set task.

3.0 EXPERIMENTAL DESIGN

The methodology for the present study is divided into two parts: (i) the elicitation of laughter from participants and (ii) the speaker identification tasks carried out by a subset of those same participants.

3.1 LAUGHTER ELICITATION SESSION

3.1.1 PARTICIPANTS

Seven females enrolled on an undergraduate course in English Language and Linguistics at the University of Huddersfield were recruited through personal connections. The group of females were recruited as they provided a social network of adequate size. We did not have access to any mixed-sex or male-only social networks, therefore the study is based solely on female voices. Participation was voluntary and written consent was obtained. They were assigned pseudonyms and will hereafter be referred to collectively as the network members. Their ages ranged from 20-21 and they were all native speakers of English who grew up in English-speaking households and originated from Yorkshire. The resulting samples were therefore rather homogeneous in terms of speaker sex, age, and geographical origin. These females were deemed to form a close social network on the basis that they had spent at least four hours per week of academic time and an unknown amount of social time together for at least two years prior to the time of this study. Though this is not a particularly empirical basis for identifying a close social network, it was the most appropriate procedure given the time and practical constraints, and is slightly more objective than Foulkes and Barron (2001:183), who based their identification of a social network of males on their personal knowledge of them and unquantified notions of the amount of time they had spent together.

In addition to these seven network members, two females enrolled on other undergraduate courses at the University of Huddersfield were also recruited through personal connections. Participation was voluntary and written consent was obtained. These 2 females were intended to serve in the speaker identification task as foils. They will be referred to as

Foil 1 and Foil 2. These two foils were selected for being unknown to the network members and for having similar sociolinguistic demographics to them. They were similar in age (21 and 25, respectively) and were native speakers of English who grew up in English-speaking homes. They also originated from Yorkshire like the network members. For the remainder of this subsection on the laughter elicitation session, the 2 foils and 7 network members will be referred to as participants when it is necessary to discuss them as a collective.

3.1.2 STIMULI

A video consisting of several film clips was compiled in *Windows Movie Maker* (Microsoft 2012) and uploaded to a computer tablet. As can be seen in Table 1 (below), 10 scenes from 5 films were chosen, totalling 374 seconds. These scenes were chosen in the hope that they would be deemed humorous by the participants and consequently elicit laughter. Of course, the extent to which the participants would find the scenes humorous and laugh, if they laughed at all, was uncertain. Accordingly, two steps were taken in an attempt to increase the chances of eliciting laughter. Firstly, all of the chosen films were contemporary productions released within the last 15 years. These were assumed to be more likely than older productions to appeal to the young participants. Of course, this was just an assumption, and it was recognised that it could be however likely, for instance, that the participants would prefer older films. Secondly, several short clips, rather than, for instance, a few long clips, were selected on the assumption that the larger the range of material, the more likely that at least some of it would be deemed humorous by the participants. 10 clips of 374 seconds in length was settled upon as a suitable amount of material for eliciting laughter. It was thought that the participants may start to lose their attention if the video was any longer. Despite these precautionary measures, however, their basis in assumptions meant that the uncertainty surrounding how much laughter would be elicited still remained. Being unable to control how much laughter would be elicited meant that participants produced wide ranging amounts of laughter, which restricted the size options of the speaker samples which were constructed from the elicited laughter for the speaker identification task (see below). However, whilst other methods, such as instructing participants to act out or produce laughter on command, may allow a greater degree of control over the amount of laughter produced, they have their own undesirable shortcomings, as noted earlier. The present study's attempt to elicit natural laughter is one methodological aspect that distinguishes it from previous studies by Yarmey (2004) and Philippon et al. (2013) which use acted laughter in speaker identification tasks.

None of the clips contained laughter in order to control for any possible effects that hearing laughter may have on the type and occurrence of laughter (Hofmann et al. 2015). Except for Mowrer et al. (1987), who use a stand-up comedy routine, previous studies that detail the contents of their elicitation material also use film scenes with no audible laughter (see Bachorowski et al. 2001; Makagon et al. 2008).

Table 1: *Movie Clips Used in the Laughter Elicitation Session Video.* (Clips are presented here in their order on the video. The movie titles, scenes, and durations are provided).

Movie	Scene(s) description	Duration (sec)
***Pitch Perfect* (2012)**	Fat Amy signs up; 'Better not'; Cynthia's Confession.	74
***Step Brothers* (2008)**	Bunk Beds; 'Even if there's a fire'.	89
***Bridesmaids* (2011)**	Bridal Shower.	57

***Mean Girls* (2004)**	Health class; 'She doesn't even go here'; Candy Cane Delivery.	58
***The Proposal* (2009)**	Campfire Song.	96
		Total: 374

3.1.3 CONDUCTING THE ELICITATION SESSION

Participants consented to being recorded during the elicitation session. They were, however, unaware that the purpose of the session was to elicit laughter. Knowing the purpose may have made the participants self-conscious and consequently resulted in unnatural laughter being elicited. Following the elicitation session, the two foils were debriefed with information concerning the purpose of the session, the intended use of the laughter data in a speaker identification task, and the project in general. The network members, however, were not debriefed with this information, as it was intended that they would later take part in the speaker identification task. Their performance in the speaker identification task might have been affected if they knew any such information prior to it. For instance, knowing in advance that they had been identified as a member of a close social network, even without explicit knowledge of who made up that network, could obviously aid them in delimiting the possible identities of the laughter samples in the speaker identification task.

Participants were recorded individually in an in-built recording booth with a hardwood floor and fabric walls. As no qualitative differences are known to arise in situations where people laugh alone or with others (Hofmann et al. 2015), there seemed to be no benefit in recording participants in groups. Besides, doing so may have given those participants who later took part in the speaker identification task, suspicions about the identities of the speakers. Figure 1 (below) diagrams the recording booth's layout. The video clip was played through headphones connected to a computer tablet placed on a table in front of the participants, who were free to control the volume of the video. A H2n Handy Zoom Recorder, mounted on a table-top tripod that was set to each participant's head height and placed in front of them at a distance of approximately 30 centimetres, was used to record the laughter. The presence of an audio recorder may have caused participants to realise that the session was designed to record some aspect of their vocal behaviour, potentially resulting in them becoming self-conscious and producing less naturalistic laughter. Thus, a video camera mounted on a tripod was also set up, but not actually activated, in an attempt to conceal the session's primary focus on audio data.

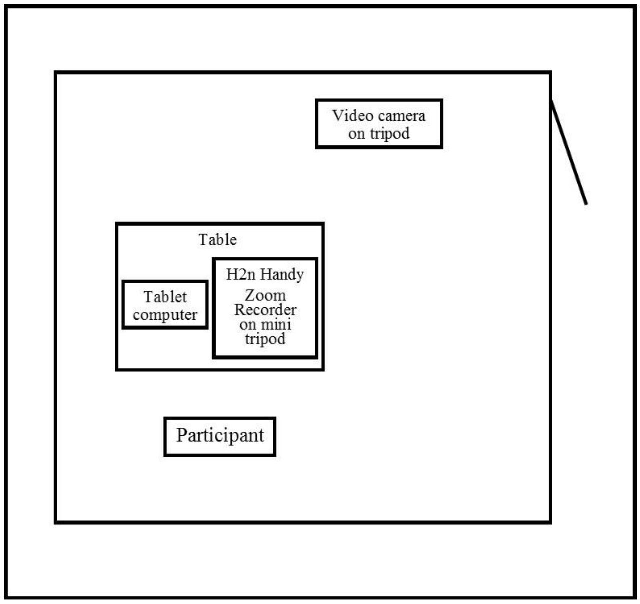

Figure 1: *Diagram of Recording Booth Layout.* (Not to scale).

3.1.4 CONSTRUCTION OF LAUGHTER SAMPLES

The recordings from the elicitation sessions were uploaded to Praat 6.0.26 (Boersma and Weenink 2017). The onset and offset of laughter was identified at the *episodic level* (Cosentino et al. 2016; see Literature Review, above) both auditorily and acoustically, and marked using Praat's TextGrids function. As the focus of this study is the audible element of laughter itself, respiration was not considered part of the laughter episodes, except for unavoidable cases when respiration occurred between two bouts that constituted one episode. Each episode was classified as being either *unvoiced*, *voiced*, or *mixed* (Bachorowski et al. 2001). For waveform and spectrographic examples of these three, taken from the present study, see Figures 2, 3 and 4. If a laughter episode had a majority of either unvoiced or voiced calls, the episode was classified as such. If a laughter episode did not appear to have a clear majority of either unvoiced or voiced calls, it was classified as a mixed laughter episode. This process of identifying and classifying laughter was a rather subjective exercise that was based on an intuitive auditory notion of what laughter sounds like in comparison to other vocalisations and sounds. Whilst this could have resulted in, for instance, coughs being mistaken for short laughs, any doubtful cases were discarded. Deciding where one episode ended and another began was similarly subjective, as such decisions were made on the auditory basis of whether a pause between two calls seemed great enough that the calls could be classed as belonging to separate episodes. Table 2 provides the results of this classification and gives the total length (in seconds) of the unvoiced, voiced, and mixed laughter episodes, as well as the net laughter for each participant.

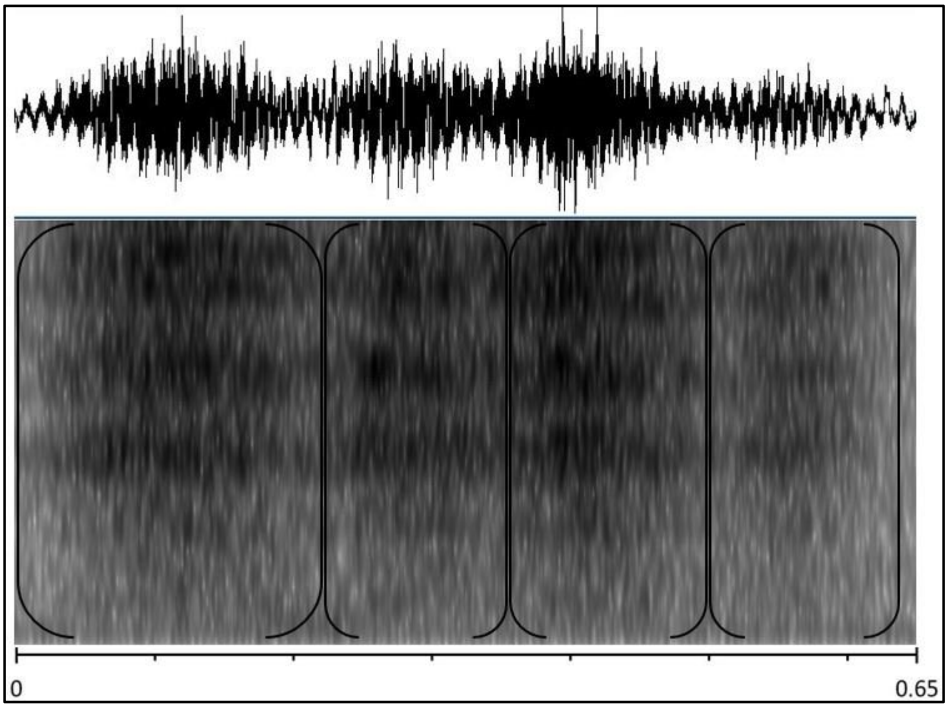

Figure 2: *Unvoiced Laughter Episode Example.* (Four calls occur in one bout. Each call is indicated by brackets in the spectrogram).

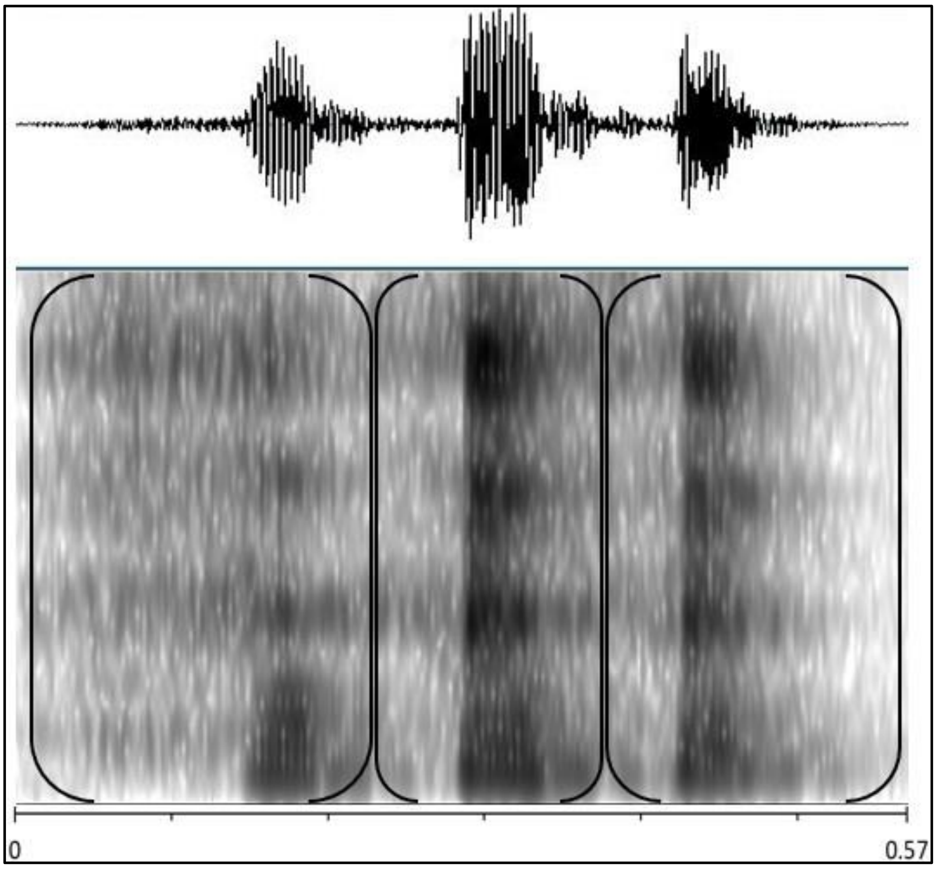

Figure 3: *Voiced Laughter Episode Example.* (Three calls occur in one bout. Each call is indicated by brackets in the spectrogram).

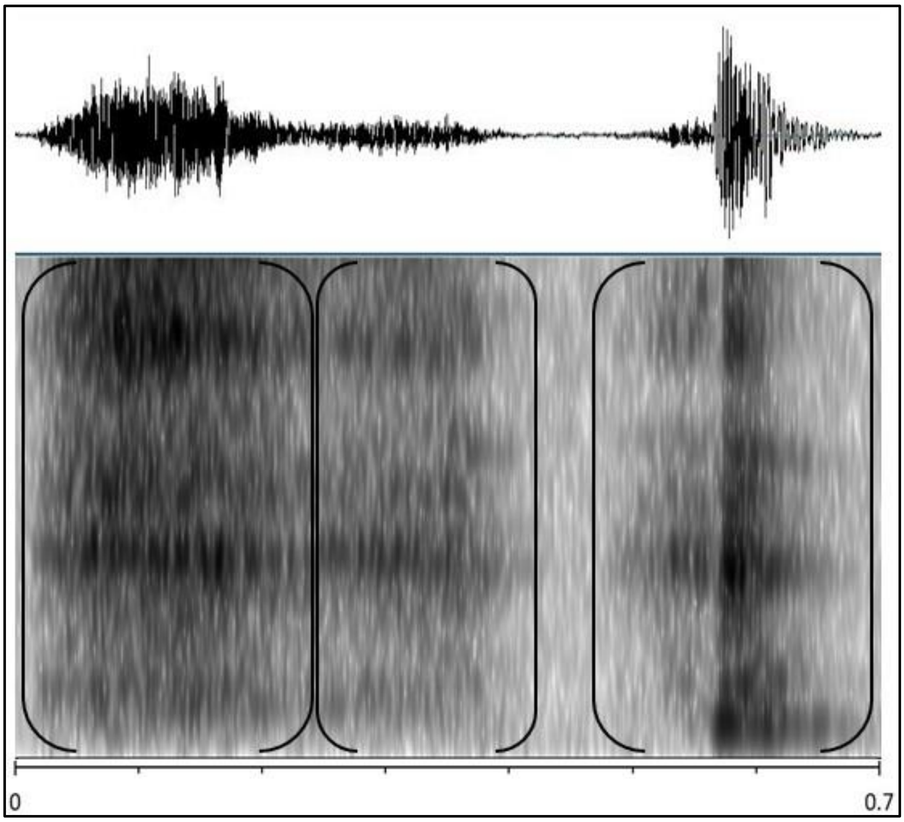

Figure 4: *Mixed Laughter Episode Example.* (Two voiceless calls precede one voiced call in one bout. Each call is indicated by brackets in the spectrogram).

Table 2: *Participants' Elicited Unvoiced, Voiced, Mixed, and Net Laughter (in seconds).*

Participant	Unvoiced Laughter	Voiced Laughter	Mixed Laughter	Net Laughter
Anne	2.45	0.73	0.82	4.0
Beth	10.51	0.23	1.77	12.51
Charlotte	3.99	1.62	0.0	5.61
Danielle	16.47	0.0	2.77	19.24
Elaine	40.92	11.47	17.49	69.88
Francine	7.85	1.75	0.28	9.88
Gertrude	1.52	3.0	1.09	5.61
Foil 1	3.23	1.19	2.99	7.41
Foil 2	4.11	1.17	1.58	6.86

Table 2 makes it clear why each sample in the identification task could not have consisted of a participant's net laughter concatenated into one file because each participant produced net laughter of widely varied lengths ranging from 4 seconds to 69.88 seconds. For example, using Anne’s net laughter of 4 seconds and Elaine’s net laughter of 69.88 seconds as samples in the identification task would have created an unfair task that gave listeners more information with which to identify Elaine. Therefore, to make a fairer task, the resulting samples all needed to be of approximately equal lengths. This length was determined by the smallest net laughter, which was produced by Anne and totalled 4.0 seconds. If greater amounts of laughter were elicited from each participant, larger samples would probably have been constructed to make the task more closely resemble the methodology of real voice line-ups (see Nolan 2003). It was also deemed necessary to create samples that were representative of the proportions of the three types of laughter episode produced by the participants for two reasons. First, this presented listeners in the identification task with all three types of laughter as opposed to just one type of laughter (cf. Philippon et al. 2013). Second, assuming that the elicited laughter was representative of the proportions of the three types of laughter episode that each speaker typically produces, those proportions may be important when identifying the speakers. For instance, listeners may associate a speaker with typically producing larger amounts of voiceless laughter relative to voiced and mixed laughter. To represent the proportions in the samples, the total amount of each of the three types of laughter episode was calculated as a percentage of the net laughter for each participant. This percentage was then used to calculate the proportions in seconds for each of the three types of laughter episode given the target sample size of 4 seconds. Table 3 provides these proportions given this target sample size.

Table 3: *Target Proportions of Unvoiced, Voiced, and Mixed Laughter given 4 seconds.*

Participant	Unvoiced Target Proportion	Voiced Target Proportion	Mixed Target Proportion
Anne	2.45	0.73	0.82
Beth	3.36	0.07	0.57
Charlotte	2.85	1.16	0.0
Danielle	3.43	0.0	0.58
Elaine	2.34	0.66	1.0
Francine	3.18	0.71	0.11
Gertrude	1.08	2.14	0.78
Foil 1	1.74	0.64	1.61
Foil 2	2.40	0.68	0.92

To construct the desired proportions, individual episodes of each corresponding type of laughter episode were then randomly selected using random numbers generated in Microsoft Excel. For comparison with Table 3, Table 4 details the actual proportions of unvoiced, voiced, and mixed laughter in each sample along with the total sample size in seconds, which ranged from 3.54 to 6.12 seconds. As can be seen in Table 4, the resulting proportions and complete samples somewhat differed in size from the target proportions. In some cases, notably the mixed laughter proportions for Beth and Danielle, this difference was particularly large. This was partly due to the random sampling process occasionally selecting laughter episodes that were extreme in terms of the typical length of that type of episode for a given speaker. It was also the result of a decision to use laughter episodes in full, rather than unnaturally cutting them for the sake of producing samples of precisely the same size.

Table 4: *Actual Proportions of Unvoiced, Voiced, and Mixed Laughter (in seconds).*

Participant	Unvoiced Actual Proportion	Voiced Actual Proportion	Mixed Actual Proportion	Total Sample Size
Anne	2.45	0.73	0.82	4.0
Beth	3.16	0.23	1.77	5.15

Charlotte	2.56	1.62	0.0	4.18
Danielle	3.36	0.0	2.77	6.12
Elaine	2.60	0.78	1.63	5.01
Francine	3.13	0.57	0.28	3.98
Gertrude	1.17	2.08	0.53	3.79
Foil 1	1.69	0.75	1.17	3.61
Foil 2	2.35	0.64	0.56	3.54

Praat 6.0.26 (Boersma and Weenink 2017) was used to concatenate the selected laughter episodes into samples. A pause of 1.5 seconds, taken from silences in the recordings, was inserted between each episode of laughter in the samples (Nolan 2003). This pause was deemed adequate to demarcate one laughter episode from another.

3. 2 SPEAKER IDENTIFICATION TASK

This subsection details the participants who took part in the speaker identification task, the preparation of the laughter samples, and the conduct of the identification task itself.

3.2.1 PARTICIPANTS

The 7 network members identified earlier took part in the speaker identification task. Participation was voluntary and written consent was again obtained. It is worth noting here that the network members, being enrolled on a linguistics course, all had undergraduate level training in phonetics and phonology, including, some experience with eliciting phonetic and phonological data. They may thus be assumed to have listening skills greater than laypeople but less than trained phoneticians.

3.2.2 TASK SET UP

The nine speaker samples obtained from the elicitation session were randomly ordered into seven separate sets, one for each network member taking part in the speaker identification task. Random orders were created to control for any ordering effects (Nolan 2003). Excel was used again to randomly assign each network member one of these sets. No network member was assigned an order that featured their own sample in first or last position, because such positions can give samples some prominence that may aid in their identification (Nolan 2003).

Each set of samples was placed into a Microsoft PowerPoint presentation. Individual samples were inserted into separate slides which had an identifying letter ranging from A to I (illustrated in Figure 5). The slides were timed for each sample to be followed by a silent five second pause in which participants could record their answers before the next sample automatically began playing. Participants heard each sample once and were not permitted to replay any of the samples. These constraints on playback and answering time were put in place because it would not have been possible to monitor them otherwise. Allowing the samples to be played at each listener’s discretion would thus have made it impossible to measure the effects of any multiple hearings on identification rates.

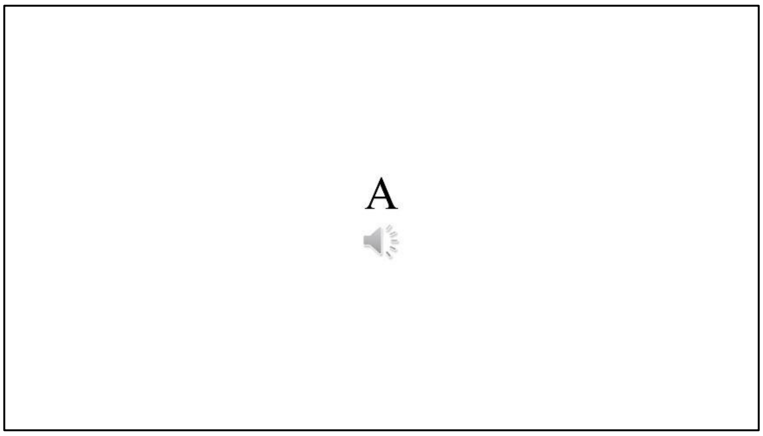

Figure 5: *Laughter Sample Presentation Slide Example.*

3.2.3 CONDUCTING THE SPEAKER IDENTIFICATION TASK

Participants took the identification task individually in the same recording booth in which the laughter elicitation sessions took place. The task was played from a laptop through close-cup headphones. Participants were told they would be taking part in a speaker identification task. This task was explained to them as one that would involve them listening to samples and attempting to identify the speaker in each case. Listeners were however also told that they did not necessarily have to provide a response. They were not told that the samples contained laughter. The task was an *open* one (Foulkes and Barron 2000). In other words, participants were not given any clues as to the identities of the speakers. Instead, they were informed that they may or may not know the speakers they were tasked with identifying. Of course, if the members of the so-called close social network were indeed as close as expected, it is likely that they knew of each other’s involvement or, if they did not know, were able to draw inferences about who may be reasonably expected to have been recruited for the samples. Thus, the notion of an open task may be more a matter of degrees of openness in this case. The participants were also told that the same sample may be played more than once, though this was never actually arranged in the sample sets. The manner in which the samples would be presented was explained to the participants and they were provided with an answer sheet labelled A-I to correspond with the presentation of the samples in the PowerPoint slides.

After completing the task, participants were debriefed about the nature of the project and asked not to share details of the session with others in case they were due to take part.

4.0 RESULTS

This section presents an overview of the results of the speaker identification task, before examining the results by listeners and speakers individually.

4.1 OVERVIEW OF RESULTS

Table 5: *Results of Speaker Identification Task.*

	Listeners						
Speakers	**Anne**	**Beth**	**Charlotte**	**Danielle**	**Elaine**	**Francine**	**Gertrude**
Anne		Beth	Beth				OIN3
Beth	Correct	Correct					OIN3

Charlotte		Beth	OIN4		Correct	OIN4	
Danielle	Anne	Beth	Elaine		Elaine	OIN6	
Elaine		Beth	OIN5		OIN2		Anne
Francine		Beth	OIN1			Correct	
Gertrude		Beth	Correct	Correct	Francine	Beth	Correct
Foil 1		Beth	Charlotte	OIN3			Danielle
Foil 2		Beth		Danielle		OIN1	

The results of the identification task overall are presented above in Table 5. Results by the network members as listeners are displayed in each column. Results by the network members and foils as speakers are displayed along each row. Cells marked 'correct' indicate that either a network member was correctly identified or a foil was correctly rejected as not being known to the listener. Blank cells indicate that no answer was provided. Cells containing a network member's name indicate that a speaker was incorrectly identified as that network member. OIN (outside identified network), followed by a number, indicates a speaker was incorrectly identified as a speaker who was outside of the identified network. Interestingly, all of the OIN speakers were enrolled on the same university course as the network members, and OIN1 and OIN2 were non-native speakers of English.

Generally, Table 5 suggests that the listeners did not perform well and that some speakers were identified more than others. Note that Beth identified herself correctly but also listed herself as the speaker in the other eight samples. Given that her own sample was in 7th position for her task, it is possible that at the point at which she correctly identified her own sample she was not making identifications by listening but was merely guessing instead. Beth's answers as a listener are therefore excluded from the following results, leaving six listeners and nine speakers. For the purposes of comparison to other studies, some results are presented with and without consideration of the answers given to the foil samples.

4.2 RESULTS BY LISTENER

The listeners' performance in this task was poor. As Table 6 shows, there was a total of 6 correct answers out of 54, 1 from each listener. Each listener correctly identified speakers 11% of the time. If the answers to the foil samples are excluded, this correct identification rate rises to 14%. 2 of these correct identifications were of the listeners (Francine and Gertrude) themselves, meaning that four of the six listeners failed to identify themselves. The other 4 correct identifications were of other network members, as no listeners correctly rejected either of the two foils. Four of the six listeners misidentified the foils at least once. Though no two samples in each task were from the same speaker, Gertrude gave the same answer twice.

The listeners gave no answer 28 times (52%) and incorrect answers 20 times (37%). If the answers to the foil samples are excluded, listeners gave no answer 21 times (50%) and incorrect answers 15 times (36%). In either case, listeners therefore tended to leave speakers unidentified more than they misidentified them. Of the incorrect answers given to both the network member and foil samples, 10 (50%) were misidentifications of speakers as other network members and 10 (50%) were misidentifications of speakers as OINs.

Table 6: *Results by Listener.*

Listener	Correct	Unidentified	Incorrect

			Network Members	OINs
Anne	1	7	1	0
Charlotte	1	2	3	3
Danielle	1	6	1	1
Elaine	1	5	2	1
Francine	1	4	1	3
Gertrude	1	4	2	2
Total	6	28	20	
%	11	52	37	

4.3 RESULTS BY SPEAKER

Some speakers appeared to be easier to identify than others (see Table 7). Three of the network members (Anne, Danielle, and Elaine) and the two foils were never correctly identified, whilst Gertrude was correctly identified the most (n=3). All speakers were misidentified and left unidentified at least once.

Table 7: *Results by Speaker.*

Speaker	Correct	Unidentified	Incorrect	
			Network Members	OINs
Anne	0	4	1	1
Beth	1	4	0	1
Charlotte	1	3	0	2
Danielle	0	2	3	1
Elaine	0	3	1	2
Francine	1	4	0	1
Gertrude	3	1	2	0
Foil 1	0	3	2	1
Foil 2	0	4	1	1
Total	6	28	20	

Table 8 (below) shows for whom each speaker was mistaken. Gertrude was misidentified as two other speakers, Beth and Francine. Beth and Francine were, however, not mistaken for Gertrude, nor were any other speakers. Danielle and Elaine were the most misidentified by number of speakers, with both being mistaken for three separate speakers. There were three instances in total where speakers were misidentified as the two non-native speakers of English (OIN1 and OIN2) by three different listeners. It is also worth noting that there were no instances of any of the female speakers being misidentified as males.

Table 8: *Misidentifications of Speakers.*

Speaker	Incorrect Answers – Misidentifications	
	Network Members	OINs
Anne	Beth (x1)	OIN3 (x1)
Beth	----	OIN3 (x1)

Charlotte	----	OIN4 (x2)
Danielle	Elaine (x2), Anne (x1)	OIN6 (x1)
Elaine	Anne (x1)	OIN2 (x1), OIN5 (x1)
Francine	----	OIN1 (x1)
Gertrude	Beth (x1), Francine (x1)	----
Foil 1	Charlotte (x1), Danielle (x1)	OIN3 (x1)
Foil 2	Danielle (x1)	OIN1 (x1)

5.0 DISCUSSION

The results of the present study are first compared with the results of a speaker identification task that used speech samples (Foulkes and Barron 2000) in order to contrast the different success rates using speech versus using laughter in voice line-ups. The results are then considered along with studies by Philippon, Randall and Cherryman (2013) and Yarmey (2004), which examined laughter using naïve speaker identification.

It is important to consider how speaker identification using laughter samples generally compares to the results of Foulkes and Barron's (2000) study, which examined the performance of members of a close social network in a speaker identification task that used samples of speech. Note that, as Foulkes and Barron (2000) do not consider listeners' answers to the foil samples in their calculations of identification rates, the results of the present study are also presented without consideration of the foils in this particular comparison. In the present study, the correct identification rate was 14% for laughter. Foulkes and Barron (2000) report a much higher correct identification rate of 68% for speech. In Foulkes and Barron (2000), only 1 of 9 listeners failed to identify themselves from a sample of their speech, while 4 of 6 listeners in the present study failed to identify themselves from a sample of their laughter. However, listeners left speakers unidentified more than they misidentified them in both the present study of laughter (50% and 36%) and the previous study of speech (22% and 10%). In terms of results by individual speakers in Foulkes and Barron (2000), all 10 network members were correctly identified at least twice, 9 were correctly identified at least 4 times, and 2 were correctly identified by all 9 listeners from samples of speech. By contrast, in the present study, none of the network members were correctly identified by all listeners, 3 were never correctly identified, 4 were correctly identified at least once, and only 1 was correctly identified 4 times. This suggests that higher identification rates may be gained from samples of speech than samples of laughter. The high degree of comparability of Foulkes and Barron's (2000) study and the present study adds some validity to this suggestion. Both were conducted in very similar ways, including the number and assumed closeness of the network members and the openness of the speaker identification tasks, though there was a difference in length between their speech samples (approximately 8-10 seconds) and the present study's laughter samples (approximately 3-6 seconds). This difference in sample size may partially explain the difference in identification rates, as the listeners presented with speech samples had more information with which to make their identifications than in the present study. Foulkes and Barron (2000) suggest that the generally good identification performance may be attributed to the speakers' relative average pitch and pitch range. Indeed, just as features arising from voiced speech appear to aid speaker identification, the next part of this discussion will suggest that voiced laughter, which was not particularly prevalent in the laughter samples of the present study, may facilitate higher rates of speaker identification than voiceless laughter.

Philippon et al. (2013) report higher identification rates than the present study. The difference between identification rates is more pronounced if the answers given to foils are included (47% compared with 11%), but a difference still exists without consideration of the foils (25% compared with 14%). These differences in identification rates might be explained by the respective studies' laughter samples being quantitatively and qualitatively different. Whilst Philippon et al. (2013) used approximately 15 second samples of purely voiced laughter, the present study's approximately 4 second samples of voiced, unvoiced, and mixed laughter contained only a small amount (ranging from 0 to approximately 2 seconds) of voiced laughter. This suggests that higher identification rates may be gained from longer samples of voiced laughter, whilst voiceless laughter may not be useful in naïve speaker identification. Indeed, in the present study, the most frequently identified speaker had by far the largest amount of voiced laughter in their sample in comparison to other speakers, as indicated in Table 9, suggesting that voiced laughter may have speaker-specific properties. The possibility that voiceless laughter may not be speaker-specific is supported by previous findings that whispered, and therefore voiceless, speech makes speaker identification difficult (see Bartle and Dellwo 2015). Additionally, following the speaker identification task in the present study, several listeners remarked that they struggled to hear any difference between the voiceless laughter in the different samples.

Yarmey's (2004) study reports an even higher identification rate of 52%. However, he does not provide any information about the voicing of the laughter samples used. This lack of information makes it difficult to interpret Yarmey's (2004) results in terms of the present study's suggestion that longer samples of voiced laughter may facilitate higher identification rates. It is also worth remembering though, that Yarmey's (2004) use of a relatively closed task may at least partly account for the higher identification rate.

Having suggested that longer samples of voiced laughter may facilitate higher identification rates, it is worth investigating whether this explains why some network members were identified more than others in the present study. Table 9 compares the size of each speaker's voiced laughter proportion (in descending order) with the frequency with which they were identified.

Table 9: *Identifications of Network Members Ranked by Proportion of Voiced Laughter.*

Speaker	Voiced Laughter Proportion	Correctly Identified
Gertrude	2.08	3
Charlotte	1.62	1
Elaine	0.78	0
Anne	0.73	0
Francine	0.57	1
Beth	0.23	1
Danielle	0.0	0

The most identified speaker, Gertrude, had the most voiced laughter in her sample, whilst Danielle, with no voiced laughter, was never identified. However, Charlotte had nearly as much voiced laughter as Gertrude, yet Charlotte was identified only as much as Francine and Beth, who had much smaller amounts of voiced laughter. Elaine & Anne in fact had larger amounts of voiced laughter than Francine and Beth but were never correctly identified. This suggests that simply exposing listeners to longer amounts of voiced laughter does not necessarily result in higher identification rates, and that additional factors may be involved. Some factors related specifically to the present study's listeners and the difficulty of the speaker identification task will be covered below. However, other properties of voiced laughter may also constitute factors affecting identification. One such property may be the fundamental frequency (hereafter, F0). F0 is reportedly important in naïve speaker identification using speech, particularly in making it easier to identify speakers with F0 mean and range values that fall outside of the norm (Foulkes and Barron 2000; Sørensen 2012; Blatchford and Foulkes 2006). Therefore, perhaps F0 is also important in naïve speaker identification of voiced laughter. Other acoustic properties, such as formant frequencies, and temporal properties, such as varied call durations, which have been cited in some of the laughter literature (see Section 2) may also be important. In addition to properties of voiced laughter, the proportions of the three types of laughter episode relative to one another could be a factor affecting identification. As mentioned Section 3.1.4, this was one of the motivations for representing the proportions of laughter types in the samples. Gertrude, for example, might be associated by her fellow network members as being someone who tends to produce the most voiced laughter, making her more easily identifiable from her sample in the identification task. Testing these possibilities against the present study's results is beyond the scope of this article. However, the possibility that F0, other properties of voiced laughter, and proportions of the different types of laughter might affect identification rates is a potential area for future research.

It was noted in the preceding paragraph that factors related to the listeners themselves may have contributed to the identification rates in the present study. Firstly, it seems unlikely that each network member was equally close to the other network members and the network as a whole. For instance, due to a personal affiliation with the network, it is known that Anne and Beth are much closer to each other than they are to the rest of the network, which may explain why Anne only correctly identified Beth and even why Anne was in fact mistaken for Beth by another listener. Secondly, whilst the time when each network member last communicated with the others prior to taking part in the speaker identification task was unknown and uncontrollable, the delay between previous exposure and testing can affect identification rates (see Clifford, Rathborn and Bull 1981). Thirdly, though the listeners' phonetic training was not to the level of an expert, the tendency of experts to outperform truly naïve listeners in speaker identification (see Schiller and Köster 1998; cf. Bartle and Dellwo 2015) may suggest that naïve listeners would be even less successful.

The listeners' judgements in relation to speaker sex and native language are worth noting. The fact that no speakers were misidentified as males might suggest that naïve listeners are sensitive to differences between male and female laughter. Differences in F0 and the first three formant frequencies in voiced laughter can be used to automatically classify speaker sex (see Bachorowski et al. 2001). Perhaps naïve listeners also attend to these features. As the speaker samples in the present study contained very little voiced laughter, maybe only a small amount of F0 and formant frequency information is needed to make judgements of speaker sex. Alternatively, additional properties of voiced laughter may contribute to making such judgements. One of the listeners who misidentified a network member as one of the non-native speakers of English from outside the network claimed that she did this because the laughter 'sounded foreign'. There are no suggestions in the literature that laughter indexes native language. However, this identifies a potential topic for future research.

The low identification rates in the present study may be partly accounted for by four potentially difficult procedural aspects of the task itself. Firstly, the listeners were not told in advance that the samples consisted of laughter and, as they were told that this was a *speaker* identification task, may have expected samples of speech. Some initial confusion or distraction from the task may have therefore occurred, particularly during the first samples. However, as only two listeners gave no answer to the first samples they were exposed to, it seems difficult to attribute possible initial confusion to the poor performance overall. Besides, listeners may have suspected a link between the previous laughter elicitation session and the speaker identification task that caused them to anticipate the involvement of laughter. Secondly, the cautionary note that the same speaker may be heard more than once may have influenced the listeners. However, insofar as the listeners' answers indicate this influence, only Beth, who was excluded earlier anyway, and Gertrude, who gave the same answer twice, appear to have been affected. Thus, this second difficulty also does not seem to account for the poor identification rates found across all the listeners. Thirdly, and perhaps more likely to have affected overall performance, were the necessary constraints on playback as well as the small speaker sample size. Finally, the presentation of an open task confused two listeners, who asked if the instruction that they may or may not know the speakers meant that they would hear celebrity speakers. To maintain control across the listeners, these two were given no clarifications in response. More generally, the relatively open task and the undisclosed inclusion of foils may have caused difficulties, and listeners may have used external knowledge or made inferences to help make identifications. Whilst misidentifications are not always the result of deductions, the fact that the numerous misidentifications of speakers outside the network were all of speakers enrolled on the same course as the network members may indeed suggest that the listeners were attempting to deduce speaker identity from a closed set that comprised their fellow classmates. As both the listeners' expectations about the content of the samples and the warning that speakers may be heard more than once appeared to affect only a few listeners, the task's openness and the constraints on playback seem the most credible explanations, at least in terms of the task's difficulty, for the listeners' poor performance overall.

6.0 IMPLICATIONS AND FUTURE RESEARCH

The results of the present study have implications for the wider forensic phonetic context, specifically voice line-ups, transcription, and forensic speaker comparison.

The speaker identification task in the present study resembled, except for the smaller speaker sample size and the constraints on playback, the methodology of a genuine voice line-up as detailed in Nolan (2003). However, many factors that are often highly variable and outside the expert's control in the forensic context, such as ear-witness stress or their familiarity with the speaker, were controlled and thus idealised in the present study's experimental context, meaning its results should be interpreted with caution. That aside, it seems that ear-witnesses would perform poorly, particularly with voiceless laughter, in a real voice line-up that used laughter. It is reassuring, however, that the present listeners tended to provide no answer rather than an incorrect one, as this is more favourable than incorrectly identifying an innocent speaker as the criminal in real cases. It is important to note that we would not advocate the use of only laughter in voice line-ups (except in exceptional cases) based on the current results and the fact that collecting adequate laughter samples from suspects would most likely be extremely challenging, given the fact that in the UK these samples are typically collected from police interviews, somewhere where you would not expect much laughter to be recorded.

As the present study's results suggest that voiced laughter may be speaker-specific, there is also the potential for voiced laughter to be used as a non-linguistic speaker discriminant in forensic speaker comparison. Though further work would be required to achieve this, initial research might investigate which properties of voiced laughter make it speaker-specific. As suggested earlier, fundamental frequency is one possible starting point. Later concerns might focus on the inter-speaker and intra-speaker variability of laughter, particularly given that, for example, F0 is known to exhibit high intra-speaker variability in speech (Braun 1995) and possibly laughter (see Hirson 1995). The potential use of such future research is emphasised by reports of a few experts having already analysed laughter without having any analytical frameworks for evaluating the discriminant power of laughter (see Gold and French 2011; Hirson 1995). Developing voiced laughter as a speaker discriminant feature may also be useful when attributing utterances to speakers in transcription. Finally, though the listeners' judgements of speaker sex and native language are less related to the present study's main concern with individual identity, further research on whether laughter indexes such speaker demographics may be useful in speaker profiling.

In conclusion, much work on the potential speaker-specificity of voiced laughter still needs to be conducted, but future research may find that voiced laughter proves a useful addition to the forensic speech scientist's analytical repertoire.

REFERENCES

Armony, J. L., Chochol, C., Fecteau, S., and Belin, P. (2007). Laugh (or cry) and you will we remembered: influence of emotional expression on memory for vocalizations. *Psychological Science* 18(12): 1027--1029. doi:10.1111/j.1467-9280.2007.02019.x

Bachorowski, J., Smoski, M. J., and Owren, M. J. (2001). The acoustic features of human laughter. *Journal of the Acoustical Society of America* 110(3): 1581--1597. doi:10.1121/1.1391244

Bartle, A., and Dellwo, V. (2015). Auditory speaker discrimination by forensic phoneticians and naive listeners in voiced and whispered speech. *The International Journal of Speech, Language and the Law* 22(2): 229--248. doi:10.1558/ijsll.v22i2.23101

Blatchford, H., and Foulkes, P. (2006). Identification of voices in shouting. *International Journal of Speech, Language and the Law* 13(2): 241--254. doi:10.1558/ijsll.2006.13.2.241

Boersma, P. and Weenink, D. (2017). *Praat: Doing Phonetics by Computer* [Computer program]. Version 6.0.26, retrieved from http://www.praat.org/ on 24/11/2016.

Braun, A. (1995). Fundamental frequency: how speaker-specific is it? In A. Braun and J. Köster (eds.) *Studies in Forensic Phonetics* 9--23. Trier: Wissenschaftlicher Verlag.

Clifford, B. R., Rathborn, H., and Bull, R. (1981). The effects of delay on voice recognition accuracy. *Law and Human Behavior* 5(2/3): 201--208. doi:10.1007/BF01044763

Cosentino, S., Sessa, S., and Takanishi, A. (2016). Quantitative laughter detection, measurement, and classification – a critical survey. *IEEE Reviews in Biomedical Engineering* 9(1): 148--162. doi:10.1109/RBME.2016.2527638

Drew, P. (2008). Conversation analysis. In J. Smith (ed.), *Qualitative Psychology: A Practical Guide to Research Methods* (2nd ed.) 133--159. Los Angeles: SAGE.

Foulkes, P., and Barron, A. (2000). Telephone speaker recognition amongst members of a close social network. *The International Journal of Speech, Language and the Law 7*(2): 180--198.

Glenn, P. J., and Holt, E. (2013). *Studies of Laughter in Interaction*. London: Bloomsbury.

Gold, E., and French, P. (2011). International practices in forensic speaker comparison. *The International Journal of Speech, Language and the Law* 18(2): 293--307. doi:10.1558/ijsll.v18i2.293

Hirson, A. (1995). Human laughter - a forensic phonetic perspective. In A. Braun and J. Köster (eds.) *Studies in Forensic Phonetics* 77--86. Trier: Wissenschaftlicher Verlag.

Hofmann, J., Platt, T., Ruch, W., Niewiadomski, R., and Urbain, J. (2015). The influence of a virtual companion on amusement when watching funny films. *Motivation and Emotion* 39(3): 434--447. doi:10.1007/s11031-014-9461-y

Holt, E. (2016). Laughter at last: playfulness and laughter in interaction. *Journal of Pragmatics* 100: 89--102

Hudenko, W. J., Stone, W., and Bachorowski, J. (2009). Laughter differs in children with autism: an acoustic analysis of laughs produced by children with and without the disorder. *Journal of Autism and Developmental Disorders* 39(10): 1392--1400. doi:10.1007/s10803-009-0752-1

Makagon, M. M., Funayama, E. S., and Owren, M. J. (2008). An acoustic analysis of laughter produced by congenitally deaf and normally hearing college students. *Journal of the Acoustical Society of America* 124(1): 472--483. doi:10.1121/1.2932088

Microsoft. (2012). *Windows Movie Maker* [Video Editing Software]. Retrieved from https://support.microsoft.com/en-gb/help/14220/windows-movie-maker-download on 24/11/2016.

Mowrer, D. E., LaPointe, L. L., and Case, J. (1987). Analysis of five acoustic correlates of laughter. *Journal of Nonverbal Behavior* 11(3): 191--199. doi:10.1007/BF00990237

Nolan, F. (2003). A recent voice parade. *The International Journal of Speech, Language and the Law* 10(2): 277--291.

Petridis, S., Martinez, B., and Pantic, M. (2013). The MAHNOB laughter database. *Image and Vision Computing* 31(2): 186--202. doi:10.1016/j.imavis.2012.08.014

Philippon, A. C., Randall, L. M., and Cherryman, J. (2013). The impact of laughter in earwitness identification performance. *Psychiatry, Psychology and Law* 20(6): 887--898. doi:10.1080/13218719.2013.768194

Provine, R. R., and Emmorey, K. (2006). Laughter among deaf signers. *Journal of Deaf Studies and Deaf Education* 11(4): 403--409. doi:10.1093/deafed/enl008

Schiller, N. O., and Köster, O. (1998). The ability of expert witnesses to identify voices: a comparison between trained and untrained listeners. *The International Journal of Speech Language and the Law* 5(1): 1--9.

Sidnell, J. (2010). *Conversation Analysis: An Introduction*. Chichester: Wiley-Blackwell.

Smoski, M., and Bachorowski, J. (2003a). Antiphonal laughter between friends and strangers. *Cognition & Emotion* 17(2): 327--340. doi:10.1080/02699930302296

Smoski, M. J., and Bachorowski, J. (2003b). Antiphonal laughter in developing friendships. *Annals of the New York Academy of Sciences*, 1000(1): 300--303. doi:10.1196/annals.1280.030

Sørensen, M. H. (2012). Voice line-ups: speakers' F0 values influence the reliability of voice recognitions. *The International Journal of Speech Language and the Law* 19(2): 145--158. doi:10.1558/ijsll.v19i2.145

Szameitat, D. P., Darwin, C. J., Szameitat, A. J., Wildgruber, D., and Alter, K. (2011). Formant characteristics of human laughter. *Journal of Voice* 25(1): 32--37. doi:10.1016/j.jvoice.2009.06.010

Vettin, J., & Todt, D. (2004). Laughter in conversation: features of cccurrence and acoustic structure. *Journal of Nonverbal Behavior* 28(2): 93--115. doi:10.1023/B:JONB.0000023654.73558.72

Yarmey, A. D. (2004). Common-sense beliefs, recognition and the identification of familiar and unfamiliar speakers from verbal and non-linguistic vocalizations. *The International Journal of Speech, Language and the Law* 11(2): 267--277. doi:10.1558/sll.2004.11.2.267