Gesture elicitation to improve DJ to audience communication

Matthew B Tindall

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Masters of Science by Research

The University of Huddersfield

July 2021

Copyright Statement

Abstract

When DJs perform they struggle to communicate their performance actions to their audience. DJs use turntables, CDJs, DJ mixers and/or computers with or without hardware DJ controllers. This equipment has small controls that are difficult to view from any distance. DJs bend over their equipment while performing small hand movements that are difficult for the audience to see. This research aims to investigate whether this visual communication problem can be solved by using full body gestures. The underlying motivation was to enhance DJ performance and the overall audience experience. Following a review of the relevant literature, this thesis begins by identifying common DJ techniques. Then gestures were elicited for each common DJ technique using the Gesture Elicitation Study (GES) methodology with the aim of creating a universally understood gesture set. The GES resulted in mainly low consensus, conflicting and inconsistent gestures which prevented an end-user gesture set from being directly produced. Therefore, three further gesture set creation studies were performed to try to create a conflict and inconsistency free gesture set. This project successfully created an end-user gesture set from the results of all four experiments. However, the inconsistencies and conflicts from these experiments suggest that that there is not a universal language that both DJs and audience members understand. Therefore, the strict GES method is deemed inappropriate for producing a DJ-audience communication focused gesture set; the author suggests adapting this methodology to involve subjective ratings to select the most suitable gestures.

Table of Contents

Deleted:

Deleted: 6

Deleted:

Deleted:

Deleted:

Deleted:

Deleted:

Deleted:

Deleted:

Deleted: 7

Deleted:

Deleted:

Deleted:

Deleted:

Deleted:

Deleted:

List of Figures

Deleted: 6

Deleted: 8

Deleted: 8

Deleted: 9

Deleted: 9

Deleted: 9

List of Tables

1 - Introduction

Within the last two decades the Disk Jockey (DJ) has risen to the height of a superstar who now appears at the centre of millions of people's social lives (Golden, 2012). The DJs popularity can be compared to the popularity that was previously received by rockstars. This shift in popularity is a testament to the heights that the DJ has reached which has seen DJs been compared to famous musicians such as Mozart and Jimmy Hendrix (Brewster & Broughton, 2006).

1.1 DJs and DJ equipment

In the most basic form DJing consists of "presenting a series of records for an audience's enjoyment" (Brewster & Broughton, 2006, p. 17). Beamish et al. (2004) defines two distinct types of DJ: the mix DJ and the scratch DJ. The mix DJ presents a constant stream of music focusing on seamless transitions and choice of material. The scratch DJ focuses on producing a range of unique noises by "pushing a record back and forth to create a scrubbing sound as the needle rubs along the groove" (Beamish et al., 2004, p. 2). Arguably the scratch DJ displays true virtuosity which can be seen as transforming the turntable from a playback device into a musical instrument (Brewster & Broughton, 2006). More recently a third type of DJ has become more common: the live DJ. This type of DJ mixes live audio from either hardware and/or software with music from playback devices. The balance between the playback of music and audio that is performed/improvised live is specific to each DJ. Having started as a scratch DJ, DJ Grandmaster Flash was one of the first DJs to add a drum machine to his setup, making him one of the first live DJs (Brewster & Broughton, 2006, p. 248). Today's live DJs may utilise a range of equipment including synthesisers, sequencers, external effects units and MIDI controllers. The majority of DJs found in clubs, whether it be headlining DJs or resident DJs, consist of mix DJs. For this reason, this research focuses on the mix DJ.

The mix DJ’s equipment tends to consist of a separate mixer and separate playback devices. Historically the playback devices were turntables with vinyl records. This remained the industry standard throughout the late 90’s until CDs gained in popularity with the general public. Over a period of time this led to CD playback devices being introduced for DJs. Pioneer released the first DJ focused CD player in October 1994 (Pioneer DJ, 2014a). However, many DJs were resistant to change until the release of the pioneer CDJ-1000 in July 2001 (Pioneer DJ, 2014a) which was the first CD player that accurately emulated the vinyl turntable with a 206mm platter. The CDJ-1000 quickly became the industry standard which could be found in the majority of DJ booths in clubs around the world up until the release of its successor, the CDJ-2000 in September 2009 (Pioneer DJ, 2014a). Figure 1 displays an annotated image of a standard Pioneer CDJ and mixer setup. Originally, commercial DJ mixers only facilitated control of volume and EQ of the playback devices. Subsequently extra functionality was added to the mixer, for example, the ability to add effects and filtering. The world’s first commercially available DJ mixer, the Pioneer DJM 500, added ‘beat effects’ to the DJ’s arsenal of tools (Pioneer DJ, 2014b). The Pioneer beat effects enabled the automatic synchronisation of time-based effects with the audio inputs through built-in BPM detection. Professional DJ Rodger Sanchez aims that the mixers built-in effects “created an additional level and a layer to my performances” (Pioneer DJ, 2014b, 1:46).

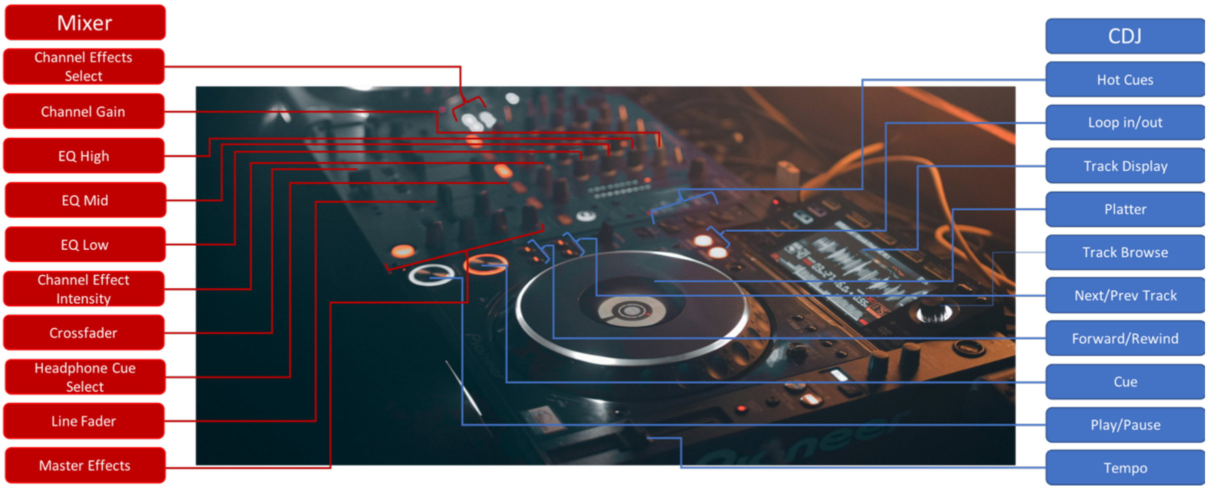


Figure 1. A standard Pioneer CDJ and mixer set up (Gurin, 2020)

Although mixers and CDJs have received a multitude of new features, such as cue points, updated effects, looping capabilities, inline filters etc… the standard technique of mixing one track into another has remained the same. Even new digital controllers, that have received wide usage due to their ease of transport and relatively cheap cost (DJ Play It, 2019; Statista, 2021), mimic traditional DJ interfaces; with a mixer at the centre of the device and two 'playback' devices at either side (Figure 2). In a similar way the visual interface for software-based DJ systems has closely emulated its hardware counterpart. (Bell et al., 2015).

Figure 2. A Traktor Kontrol S4Mk2 'all in one' MIDI DJ controller (Charlieekelly, 2013)

1.2 DJ-Audience Communication

Traditional DJ equipment results in interactions that hinder visual communication between the DJ and the audience. Mix DJs tend to perform small movements when interacting with the small dials and buttons of their hardware. These hand movements are often hard to discern, even when stood in close proximity to the DJ. This issue is compounded in medium and large size venues where it is impossible to see the DJ's interactions with their equipment. When performing small hand movements the DJ could be controlling any aspect of the sonic output and the audience would be unaware of what was happening. In contrast to this, traditional musical instruments such as the guitar, violin and drums afford much better visual communication from the performer to the audience. For example, when a drummer hits a drum, the audience clearly relates the visual action to the sonic outcome. The bowing of an orchestral string instrument and the strumming of a guitar both provide similar visual cues. These instruments have a strong action-sound coupling. Jensenius defines action-sound couplings as "relationships between actions and sounds where there is a mechanical and acoustical coupling between the action and the sound" (2007, p. 21).

Evidence that DJs are struggling to communicate with their audiences can be seen in the strategies that DJs employ in order to try to communicate how their actions relate to the sonic output. DJs often use exaggerated hand movements to make their movements more noticeable (Gates et al., 2006). For example, when some DJs move a fader on their equipment they throw their hand in the air as if receiving an electric shock to emphasis the fact that they have moved the fader. DJs also perform specific gestures such as the 'rewind' gesture in which a DJ will spin his finger round in the air to signify spinning back a record (Fintoni, 2015), or the famous 'Jesus pose' (Golden, 2013) which signifies the impending arrival of the 'drop' (the drop is where the main rhythmical elements return after having been absent in the preceding breakdown). Some DJs take gestures to the extreme by picking up their equipment and brandishing it to the crowd in order to emphasise that they are controlling the change in the sound (DanceTelevision, 2013, 36:05).

1.3 Gestures

Typically modern desktop computers still use a keyboard and mouse. However, modern devices, such as smart phones and laptops, feature gestural control via touchscreen/trackpad. Kirtenbach and Hulteen (1990, p. 310) define gestures as:

> *"… a motion of the body that contains information. Waving goodbye is a gesture. Pressing a key on a keyboard is not a gesture because the motion of a finger on its way to hitting a key is neither observed nor significant. All that matters is which key was pressed."*

Gestures afford a more natural and intuitive means of interaction (Muser, 2015). Technological advances have resulted in the widespread use of gestural interactions with electronic devices in most peoples' daily lives. An example of this can be found in modern smart phones that lack a hardware home screen button where gestures are often associated with specific tasks regardless of the operating system: tap to select, pinch to zoom , swipe up to return to the home screen, swipe up and hold to view running applications, swipe from the right to go back. More recent technological breakthroughs such as the Microsoft Kinect and the Leap Motion controller have enabled mid-air gestural control.

The NIME community (NIME, 2021) has seen a significant amount of research that focuses on designing and developing new technologies that afford the translation of movement (gestures) into sound with the aim of adding a greater degree of expression to performances. The NIME community focuses on using gestures to control musical instruments as opposed to DJ processes. Typically these research projects have gesture sets that are designed for a single performer/performance and not for wider universal use. In contrast, Imogen Heap’s MiMu gloves are an exception, as these have now been developed into a commercial product (Mimu, 2021).

1.4 Gesture Elicitation

Traditionally gesture sets have been created by teams of designers without reference to the end user. Wobbrock et al. state that this is because “designers may organize their gestures in a principled, logical fashion, user behavior is rarely so systematic” (2009, p. 1). Instead of building a DJ gesture-based system that uses gestures which were designed by a small team of designers, this research adopts a Gesture Elicitation Study (GES) (Wobbrock et al., 2009). GES were formed around Norman’s User Centred Design (UCD) approach which puts the needs of the user first and foremost (Norman, 2013). GES is a methodology used to discover gestures from a set of stakeholders. This methodology was developed in an attempt to address the above issue. Human Computer Interaction (HCI) literature states that elicited gestures are more memorable (Nacenta et al., 2013), intuitive (Ali et al., 2019), learnable and comfortable (Wu & Wang, 2013). As far as the author is aware the GES has not been utilised to develop a DJ gesture set.

1.5 Aims

The main aim of this research is to elicit a set of gestures that better communicate to the audience what the DJ is doing to the sound.

In order to achieve this, the following need to be addressed/considered:

1. Define common DJ techniques/tasks.
2. Identify which tasks need communicating to the audience through the analysis of common DJ techniques.
3. Elicit gestures for the identified tasks.
4. Validate the elicited gestures.
5. Assess whether gesture elicitation is appropriate and successful for this domain.

1.6 Structure of Thesis

Section 2 presents a review of the relevant literature, and considers: research into DJ interfaces, DJ-audience communication, gestures in music and live performance and approaches to gesture set development. Section 3 investigates the identification of DJ techniques. Then the main experiment, the Gesture Elicitation Study (GES), is detailed in section 4 in which end-user gestures are elicited, grouped and then analysed. Section 5-7 details the gesture set creation experiments. Then section 8 presents a discussion. Finally the work is concluded in section 9.

2 - Literature Review

2.1 Introduction

The following section will begin by defining the methodologies that can be used to identify common DJ techniques. Then the attempts the DJ makes to try to communicate with the audience are analysed; with an emphasis on the gestures DJs are currently performing. Then the use of gestures in musical performance and in music production are detailed; during this section the research into gesturally controlled DJ interfaces is reviewed. Finally gesture set design and evaluation methodologies are detailed, with a section focusing on the Gesture Elicitation Study (GES) methodology.

2.2 Methodologies used to Identify DJ techniques

The first step in performing a GES is to identify the tasks that the gestures execute; these are often referred to as referents. During traditional HCI studies the referents tend to be made up of primitive computer tasks such as copy and paste, select icon, zoom in/out etc… The GES participants are then presented with videos which show each referent and are then instructed to provide a suitable gesture (Wobbrock et al., 2009). In the interest of the GES performed during this thesis, common DJ techniques must be defined and then audio samples created that represent the common DJ techniques which are then played to the participants in a GES. During this section the different methods of analysing live performances are detailed with the aim of identifying the most effective method of determining common DJ techniques.

A number of methodologies could be utilised to ascertain common DJ techniques, which range from: autoethnographic, ethnographic, to a review of academic literature.

The autoethnographic methodology consists of self-study in order to explore personal experiences. Initially this form of study may appear academically unsuitable due to its subjective nature. However, a number of scholars state the importance of integrating personal experience within academia (Chang, 2008; Ellis, 2011; Anthony, 2017). Anthony (2017) presents an example of a performance based autoethnographic study in which the author video records himself whilst he mixes music in his studio with the aim of analysing his music production process with a focus on the techniques that he uses when interacting with his equipment. Anthony found that he has adopted an embodiment of the music approach; this sees Anthony moving to the music as a live performer would when performing in front of an audience. Anthony also found that his mixing process followed a more musical approach in which he utilised the mixing desk as an instrument as opposed to the modern practice of using a mouse to interact with his DAW. Anthony then discusses the importance of the experience with the equipment one uses, stating that a way of turning mixing into a performance lays in the engineer's ability to work freely with their equipment; almost acting as if the equipment is an extension of the self. These key insights would have been otherwise hard to establish whilst using ethnographic methodology. However, a negative aspect of the autoethnographic approach is the awareness that the session is being recorded and thus removing the user from the state of 'flow' (Csikszentmihalyi, 2002). Anthony states that the video examples in his study were chosen with a specific emphasis on being in the state of flow, or as Anthony states "off in my own little world mixing". Furthermore, due to the subjective nature of the autoethnographic methodology, one would be unable to define the impact that this bias has on the observations.

When applying an autoethnographic methodology to define common DJ techniques, the author would film himself DJing and then the resultant footage would be analysed utilising an inductive analysis method. There are two negatives of the autoethnographic method when applied to define common DJ techniques. Firstly, the aim of this experiment is to find common techniques, which implies that the techniques are used by most DJs. However, the autoethnographic approach focuses on a single subject, which could result in an inaccurate representation of common DJ techniques. Secondly, the author does not have access to the industry standard equipment. Therefore, one could claim that the analysis does not accurately reflect the tasks that most professional DJs are performing; as the authors DJ equipment may encourage a different style of interaction.

The ethnographic approach is similar to the autoethnographic methodology however as opposed to the analysis of one's own behaviours, the researcher analyses either single subjects or communities of people. The ethnographic methodology enables "the exploration of context through observation and presents opportunities to ask questions and discuss motivations and reasoning" (McGrath et al., 2016, p. 1). This methodology has been applied to a wide variety of areas across the music discipline, which include: the analysis of the DJ (Gates et al., 2006; Ahmed et al., 2012), the analysis of the EDM scene/festivals (Kavanaugh & Anderson, 2008; Mariano Pina, 2015) and the analysis of the workflow of music production (McGrath et al., 2016; McGarry et al., 2017).

McGrath et al. (2016) studied two producers who collaborated on the production of a grime EP. The study details the workflow of the producers who were observed in their working environment. McGrath et al. successfully defined the producers workflow. Information was gathered through informal interviews which took place on a weekly basis. The authors also had access to the private correspondence between the two producers. Having access to multiple sources of data displays the benefits of performing ethnographic research; as such a wide amount of coverage guarantees all information is taken into consideration. On the other hand, the sheer amount of data and resultant time taken to collect and analyse such data may be excessive for certain studies.

Utilising a comprehensive ethnographic method for the application of defining common DJ techniques would entail the author filming DJs performing in club settings. The DJs would be local and not famous ‘headline’ DJs, as such DJs are unlikely to give their permission to film them. The resultant footage would then be analysed using an inductive analysis. Initially, this methodology appears to be acceptable however there a number of negative aspects. A major negative of this study is the technical aspects of filming a DJ in a club environment. It is often the case that clubs are extremely dark which would require either a number of lights to illuminate the DJs equipment, which most DJs would not permit, or specialist camera equipment which the author would not have access to. Another negative aspect of such a methodology is the amount of time that filming numerous DJs would take.

In contrast, the ethnographic study could also be performed utilising third party footage. This style of ethnography does not require the in-depth study of people to the point of tracking personal communications as performed in McGrath et al. (2016). During this method, the video footage of DJs would be acquired from a third party (such as YouTube or Vimeo). There are a number of positive aspects of using this methodology when defining common DJ techniques. Firstly, the quality of the DJ and footage are guaranteed due to the significant amount of videos which are hosted by channels that are dedicated to hosting video footage of professional DJ performances (Boiler Room, n.d.; YouTube, n.d.; DanceTelevision, n.d.). Secondly, this method removes the need to perform any video recordings which saves time. Thirdly, the resultant gesture set from this project are aimed at professional DJs, therefore it seems logical that the techniques would be derived from such individuals. Finally, the DJs that feature in these videos all have access to the industry standard equipment which would result in an analysis that is representative of most professional DJs.

The final methodology consists of performing a literature review of the relevant academic literature. The various sources would then be compared with the aim of defining commonalities among the DJ techniques. There a number of methodological approaches to literature reviews such as the Systematic Literature Review (SLR) and the rapid review. The SLR follows an evidence based approach in which a formalised search for all relevant literature (evidence) is performed. Typical SLR collect results relating to a specific research question. Then, if there have been a sufficient number of studies performed, a meta-analysis can be performed to cross compare results (Kitchenham, 2010). The rapid review follows a similar approach to the SLR, yet a number of the processes are either simplified or omitted in order to complete the review in a shorter time frame. The major negative of a review based approach is the reliance on the existence of research of the topic in question. As far as the author is aware, this is the case with DJ literature; as the majority of studies appear to focus on the scratch DJ which neglects the tasks performed by the mix DJ.

2.2.1 Video Analysis

The autoethnographic and ethnographic methods require the analysis of video footage which tends to be performed following an inductive approach. Inductive analysis is a method of qualitative data analysis which aims to reduce information into a condensed format to develop a "… model or theory about the underlying structure of experiences or processes which are evident in the raw data." (Thomas, 2003, p. 1). This methodology consists of developing themes throughout the analysis which contrasts to the hypothetico-deductive methodology where a theoretical structure is defined prior to the analysis.

There are a number of popular tools that are used to perform an inductive analysis. The options range from Microsoft Excel, annotating by hand to 'Qualitative Data Analysis Software' (QDAS).

QDAS:

> *"… facilitates efficient management of qualitative and mixed methods data through a variety of tools to organize and keep track of multiple data sources and types and of the ideas flowing from those data."* (Bazeley, *2018, p. 2).*

QDAS offers the user the means to code videos in which markers (nodes) can be set at any given moment of a video to signify an event. Nodes follow a hierarchical structure that consist of parent and child nodes. An example of this can be seen in Figure 3 where the parent is 'fader' and the child is 'Up', 'Down' and 'Varying' etc... The method of 'coding' is often utilised to perform inductive analysis.

Fader	10	560
BottomToTop	2	9
Chop	1	9
Down	10	167
kill	10	56
Up	10	262
Varying	6	57

Figure 3. The hierarchal structure of a 'Fader' node. The parent appears at the top and then all child nodes are displayed below

There are several different QDAS packages available, the most prominent packages being: NVivo (NVivo, 2021), MAXQDA (MAXQDA, 2021) and ATLAS.ti (ATLAS.ti, 2021). Each package offers the video coding functionality mentioned above, however the institution in which this project was carried out offered licensing for NVivo, therefore NVivo was utilised during this project.

2.3 DJ-audience communication

DJs and audience members communicate with each other during performances.

Audience-DJ communication has been considered by several different studies which focus on three different aspects: ethnographic analysis of DJs, automated DJ systems and audience-centric DJ systems. Gates et al. (2006) and Ahmed et al. (2012) performed ethnographic studies that both state that the DJs track selection is heavily influenced by the audiences communication. This communication ranges from cheering, clapping and whistling (which are a means of complementing the DJ), to more subtle behaviours such as analysing body language for when people become tired, thirsty or bored (Ahmed et al., 2012). Feldmeier (2003) and Cliff (2006) created automated mixing tools that would potentially replace the DJ. Both research projects used sensors to influence the mixes that their applications outputted. Through monitoring the audiences' movements, both papers highlight an interest in audience-DJ communication. Similarly Graña (2020) built a audience-centric system which allows users to vote for the next song during DJ performances that are live streamed via a purpose built application.

DJ-audience communication is the focus of this project. DJs attempt to communicate with their audiences using several different methods. These methods include performing over exaggerated movements when interacting with DJ equipment, performing specific movements/gestures that signify changes in the sonic output and angling DJ equipment in the direction of the audience to afford better visibility of the interactions with the equipment. These methods can be seen to be a result of the interactions forced by commercial DJ equipment. Yet little research has been performed that analyses the DJs actions and more specifically how these actions are used as a means of communicating with the audience.

As well as considering audience-DJ communication, Gates et al. (2006) also undertook an ethnographic study of DJ-audience communication. Eleven DJs from different cities were interviewed. The study was performed in three stages. Firstly, the primary author visited numerous nightclubs where she performed exploratory research. The DJs were then given a quantitative survey regarding "DJ awareness and interaction with nightclub audiences, from an HCI perspective." (Gates et al., 2006, p. 4). After the survey data was analysed, the DJs were then interviewed regarding their answers. One of the participants stated that the way that the DJ interacts with the crowd is equally as important as the sonic output. This reinforces the importance of DJ-audience communication.

However, it is often the case that the communication is hindered by the tools that a DJ uses. This can be viewed as a result of a degradation of control intimacy (Moore, 1988). Control intimacy, as introduced by Moore, is "… how the richness and nuance of a performer's movement translates into the musical output of an instrument." (Jack et al., 2018, p. 1). Control intimacy is similar to 'transparency'; which "provides an indication of the psychophysiological distance, in the minds of the player and the audience, between the input and output of a device mapping." (Fels et al., 2002, p. 1). The lower the transparency, the more the control intimacy has been degraded.

Moore claims that the degradation of control intimacy arose with the modern MIDI controller, which contains small dials, buttons and knobs that are similar to the controls found on DJ equipment. Therefore, the DJ-audience communication problem is inherent within DJing regardless of what hardware the DJ performs with. One might argue that with vinyl DJs there is a slightly higher degree of DJ-audience communication due to the larger movements that this format entails (Figure 4); vinyl DJs are forced to cue and search by placing their hand onto the record and moving it backwards and forwards, on the other hand CDJs and DJ MIDI controllers provide the user with dedicated hardware controls such as buttons and touch scripts that result in smaller movements. However, when DJing with vinyl the performer still utilises a mixer, which houses small dials, buttons and faders. Therefore, this type of interaction is similar to that of a more modern digital set up. It is also important to note that some movements when interacting directly with the turntable are small: changing the tempo is performed via a small pitch fader, playing the record is performed by pressing the play button, temporally slowing the record down by pinching the spindle. Furthermore, the author believes that Moore's (1988) statement could be taken one step further when applied to DJing - poor visibility and expression are inherent traits that have always been present within DJ performance regardless of the format the DJ utilises.

Figure 4. DJ Schiller holding up a vinyl record. This displays the visual advantage of using a tangible format as the audience will be able to physically see what track is being played next

This poor communication resulted in DJs performing exaggerated hand movements when interacting with their equipment. This phenomenon can be seen from DJs across all genres. For example, when Trance DJ Armin Van Buren interacts with the knobs and faders on the mixer he jolts his hands upwards (DJ Mag, 2020, 4:29). House/Techno/Dubstep DJ Skream performs similar movements to Armin Van Buren throughout his DJ set (Boiler Room, 2013b). Finally, Drum and Bass (D&B) DJ Andy C jolts his hand up and tilts his whole body backwards when interacting with the knobs on the mixer (bantonblud, 2009, 4:28). Additionally, Gates et al. (2006) found that around half of the DJs in their study claimed that they exaggerate hand movements when interacting with their equipment to display what they were doing more clearly to the audience. This dramatized movement is not always seen in a negative light as one of the DJs claim that "it looks better" but as can be seen by viewing several forum posts (fbonito, 2011; sarasin, 2011; polybius, 2012; G0LDI_L0CKS, 2018), there is a mixed audience sentiment towards this type of movement. BrineWR71 (2018) and djdadi (2018) see such movements as a negative, claiming that DJs use them in order to appear to be performing more complex tasks than they actually are. On the other hand do_not_engage (2018) claims that such movements are a necessary part of DJ performance. Regardless of how people view this phenomenon, it is clear, as Gates et al. (2006) state that these movements are performed so that the audience is aware that the DJ is performing a task. This indicates that the lack of control-intimacy, and thus poor communication, is inherent in DJ performance.

Certain DJs also perform specific poses/movements to signify changes in the sonic output. The most famous example of these is the ‘Jesus’ pose (Golden, 2013). Figure 5 shows two examples of DJs performing the Jesus pose.

Figure 5. DJ Tiesto (Fonseca, 2010) and Corey Soljan (Apel, 2015) performing the Jesus pose, from left to right respectively

This pose is performed before the drop and tends to be held for a number of bars. Although the origin of this pose is unknown and thus the original context cannot be identified, one cannot help but feel that this draws upon the mental model of the rise of the DJ as Holy figure (Brewster & Broughton, 2006), in which his disciples (the audience), come to pay him homage at church (the club). Till (2010) supports this metaphor in a more general context of popular music. However, the origin and the motivation behind this gesture go beyond the scope of this project, as the action-sound relationship is what takes precedence during this discussion. In relation to this, the DJ is communicating to their audience that there is a going to be a drop and due to its widespread use and thus universal understanding, the majority of nightclub patrons will be thrown into anticipation when seeing a DJ perform this movement.

Another popular gesture is the ‘rewind’ motion which sees either the audience or the DJ spin their finger round in the air to mimic a record being wound backwards. This gesture is used to signal the track being restarted. The rewind gesture originated in the dance halls of Jamaica (Fintoni, 2015) and has since spread across the world. There is lots of controversy around this gesture as it is often used gratuitously by the DJ, Fintoni claims that it can “interrupt the flow of the music or seem to be a mere celebration of the performer’s musical ego, an attempt at trying to fake excitement” (2015). Regardless of how DJ and audience members feel about this gesture, it demonstrates that there can be a gesture vocabulary that both DJ and audience can understand, and thus utilise to aid in an enhanced audience experience.

Some DJs take exaggerated hand movements and large poses one step further by brandishing their equipment to the crowd so that the audience can clearly see what they are doing. A prime example of this behaviour can be found in any set performed by the artist ‘KiNK’ who is famed for his performances in which he picks his controllers up and tilts them towards the audience while performing actions (DanceTelevision, 2013, 36:05). This highlights the existence of the DJ-audience communication problem through the measures that some DJs are performing in order to communicate how their actions link to the sonic output.

The gestures/poses that have been mentioned throughout this section show the DJ is attempting to communicate with the audience. The rewind gesture and Jesus pose suggest that there is, at least to some extent, an existing gesture vocabulary that both DJ and audience members can understand. This leads one to consider whether other gestures could similarly be utilised to control DJ parameters. Due to the use of the body, and resultant large movement, the audience will be able to see what the DJ is doing with ease, and they will also be able to understand what the DJ is doing as opposed to watching the DJ bend over their equipment while moving small knobs and pressing small buttons. In comparison to standard hardware interactions, gestures also afford greater expression (Dobrian & Koppelman, 2006; Paine, 2009). These findings lead on to a pertinent question: are there other innate gestures/poses that both a DJ and audience members can understand which could be utilised to control DJ parameters? If so, what methodologies could be used to identify such gestures/poses? Before investigating gesture set design methodologies the use of gestures in music performance and production are reviewed.

This thesis uses the term DJ-audience communication but as has been mentioned in this section other work has referred to this as 'transparency' and 'control intimacy'.

2.4 Gestures in Music Performance and Production

There are a number of relevant areas that are closely linked to DJ performance and gestural interaction with music. Conductors have been using gestures for over two centuries (Galkin, 1988). DJs and conductors both control the playback of pre-composed music. DJs utilise dedicated hardware playback devices whereas conductors command musicians who in turn playback music. Another key figure in the music industry that has a number of similarities to the DJ is the music producer. The music producer manipulates audio through specialist equipment; often treating the equipment as an instrument (Anthony, 2017). This is similar to the way in which DJs manipulate music who also tend to treat their equipment as instruments (Webber, 2018). In addition, a number of live performers have created sensing devices which trigger and manipulate sound in a live environment. A gesturally controlled DJ interface would afford DJs to perform in a similar way.

This section begins by reviewing the conductor. The research that has been performed into controlling music production software with gestures is then detailed. Then sensing devices that produce and manipulate sound in live environments are discussed. The section is concluded with a review of gesturally controlled DJ interfaces.

2.4.1 Conductors

Conductors in their simplest form are a “leader of a musical ensemble” (Miller, 2012, p. 13). Conductors utilise gestures to direct orchestral performers. It is also important to note that their gestures not only affect the performers that they are conducting, but they also affect the viewers perceptions of the overall performance (Kumar & Morrison, 2016); here the audience receives visual cues from the conductor. This is analogous to the over exaggerated hand movements (gestures) that DJs perform in order to communicate with the audience, as both DJs and conductors gestures have a direct impact on how audience members perceive performances. Whilst currently DJs gestures don’t control the music, there is no reason gestures couldn’t be adopted for control. In this sense there are close parallels between the DJ and conductor.

Gestures are an intrinsic element of a conductor’s role. From the large swooping movements of their arms to the less significant, but by no means less important, facial gestures (luck et al. , 2006). Price (2011) performed a study that investigated how conductors affected people’s perception of orchestra performances. During the study music students scored video recordings of seven different conductors directing an orchestra for musically identical performances. Even though the performances were the same, they found that the “… music is perceived as different, depending on the conductor visual information accompanying it.” (Price, 2011, p. 10). This was linked to the expressivity of the conductors as Price concludes that the higher the level of expressivity used by the conductor, the higher the performance is rated by the audience. Madsen et al. (2007) further supports this finding by performing a study where a group of listeners were given several orchestra recordings (without the visuals) that were led by different conductors. The listeners could not discern any differences when listening to the audio without visuals.

Kumar & Morrison (2016) investigated how specific movements of the conductors affect the audience members perception of the performance. Musicians rated the video recordings of conductors leading orchestras for articulation, rhythm, style, and phrasing. The results state that “Listeners appear to be sensitive to the manner in which a conductor’s gesture delineates musical lines, particularly as an indication of overall articulation and style.” (Kumar & Morrison, 2016, p. 1). This further reiterates the importance of the conductors movements and highlights the way that their movements influence people’s perception of the overall performance.

Similar studies have been performed for a number of different musicians, ranging from singers, pianists to clarinettists (Thompson & Russo, 2007; Davidson, 2012; Thompson &

Luck, 2012). They state there is a visual dominance in such performances; especially when relating to expressivity and emotion. This highlights the multimodality of the listening experience. If the audience cannot see or understand what a DJ is doing, then one could argue that the DJs performance would result in an inferior perceived experience in comparison to a gesturally controlled DJ performance.

These observations display the importance of the visual field during musical performances. The presence of a conductor can have a large impact on an audience members perception of an orchestral performance. The more expression the conductor uses, the higher the subjective rating of the overall orchestra performance (Price, 2011). The fact that a figure stood at the front of a group of musicians can affect the listeners' perception of the music has major implications for live music. When applying these findings to a DJ performance, the relevance of the DJs movements and the expressivity of those movements are highlighted. This results in the belief that gestural control for DJs, which by nature would allow the DJs to express themselves to a greater degree in comparison to traditional DJ hardware, would result in a better perceived performance.

2.4.2 Gestures for music production

Whilst it is not a live performance to an audience, innately music production is similar to DJing as it involves the user carrying out similar audio processing by interacting with a range of technology.

There have been a number of efforts to design gestures to control Digital Audio Workstations (DAW). Balin and Loviscach (2011) performed a study with the aim of creating a touch gesture set to control a DAW. Touch gestures are gestures which are inputted via dedicated hardware sensors such as a trackpad or touchscreen. A gesture mapping activity was undertaken, in which touch gestures and DAW based tasks (referents) were presented to users via a web-based survey. The users were asked to map one of 30 gestures to one of 22 referents. This resulted in eight of the original gestures being excluded from the results; the method of excluding a number of the gestures from the resultant gesture set is seen as a positive technique as it allows the users to omit the gestures that they deem unsuitable. The referents were chosen by the authors and the development of the gestures was not described. The referents consisted of primitive DAW operations such as: go forward one bar, go backward one bar, select all and jump to end etc… The gestures consisted of: draw a line upwards/downwards, draw a line to left/right, draw circle, draw a triangle etc… The gestures that received the most agreement/consensus were the stroke up and down gestures to control intensity, the stroke left and right gestures to jump backwards/forwards bars and the metronome gesture (a triangle gesture with a finger) that (de)activated the metronome. The results of this study were not validated. Validation is the process of ensuring consistency and accuracy of results, often determined through another round of subjective testing in which the popular gestures are compared to benchmarks. The use of the stroke up/down gesture may be able to be transferred to a full body gesture in this research. Given the success of the metronome gesture this indicates that symbolic gestures, which are gestures that "visually depicts a symbol" (Wobbrock et al., 2009, p. 4), may also be useful for DJ gestures.

Lech and Kostek (2013) built a mid-air gesturally controlled system that operates a DAW. The gestures interfaced with a purpose built application that sent MIDI messages to any DAW. The visual interface consisted of circles which represented each track. The size of the circle controls the dynamics, the horizontal position of the circle controls the panorama and the vertical position of the circle represents the gain of the selected EQ band. An example of a gestural interaction of this system is extending the index and the middle finger into a 'peace' symbol to control the transport functions such as play/pause. The design methodology of the gesture set was not detailed. Mid-air gesture sets that first instruct the user to strike a pose and then perform an action are often arbitrary which result in gesture sets that have poor discoverability and memorability producing frustrating interactions (Norman, 2013).

The system was tested on 10 experienced mix engineers who were asked to perform five different mixes with specific aspects of the application disabled during each of the five mixes. The different application mixing modes are displayed in Table 1. The different mixes were then compared to define the efficacy of each application mode. After the mixes were performed, the participants were asked to subjectively rate the interface.

Table 1. Lech and Kostek's (2013) application mixing modes

Without Visuals	**With Visuals**
1. Mixing using gestures to control the engineered system.	2. Mixing using gestures to control the engineered system.
3. Mixing using mouse and keyboard to control the engineered system.	4. Mixing using mouse and keyboard to control the engineered system.
	5. Mixing in a DAW using a keyboard, mouse and a MIDI controller for audio parameter adjustments.

Regardless of the application mode, all of the users rated the overall interface highly. One of the participants claimed that using gestures enabled him to focus on the sound better however the gesture set was rated poorly by the majority of the participants. The authors claim that through observation they found two factors that resulted in the low grading of the gestures. Firstly, the design of the gesture set did not allow the user to rest their hands which resulted in fatigue and secondly the time it took for the system to recognise the gestures. However, the authors claim that using hand gesture "… interaction in sound mixing produces mixes that are not worse regarding aesthetic value than the ones obtained using DAW software handled by a mouse, keyboard, and MIDI controller." (Lech & Kostek, 2013, p. 11). This statement demonstrates that mid-air gestures are a suitable means of controlling audio processing. Due to the close link with DJing these findings could be transferred to the DJ specific domain.

Ratcliffe (2014) and Wakefield et al. (2017) created Leap Motion (Leap Motion, n.d.) controlled gestural interfaces that take advantage of the 'stage metaphor'. The stage metaphor displays audio channels through nodes in a box, or stage, similar to the visual interface described in Lech and Kostek (2013). The horizontal axis controls pan and the vertical axis controls the level.

Through a self-evaluation Ratcliffe (2014) states that his system works and is usable however it lacks commonly expected features of a DAW such as multiple tracks and effects. The gestures that are used in this system are not discussed in detail, therefore the author assumes that the only gestural interactions are that of interacting with the nodes. Wakefield et al. (2017) built LAMI, a standalone web application that contains many of the features commonly found in a DAW including: EQ, sends and multiple tracks. A number of different gestures are used to interact with the system which consist of the same interaction style as Lech and Kostek (2013), where the user must first strike a pose and then perform a movement. Users can manipulate volume, pan and both of the sends simultaneously. This form of control is similar to multi-mapping which enables the users to control a number of parameters via one control/movement. During the user evaluation of LAMI, the authors found that the participants would have preferred to control parameters individually instead of in a multi-mapped fashion. Seven out of eleven of the participants, who were all experienced in mix engineering, stated that they enjoyed using the controller, however some of the more complex gestures caused frustration.

In the aforementioned papers the origin of the gestures were either not discussed or designed by the authors. In contrast Berndt et al. (2016) performed an open house elicitation event in which the author aimed to identify the most suitable gestures to control a DAW. Prior to the event a prototype application was developed with the aim of allowing users to become accustomed to gestural input via a Leap Motion sensor. Users were instructed to interact with the system whose axes were pre-mapped with the following musical features:

- Tempo – depth of hand.
- Dynamics – vertical position of hand.
- Articulation – Opening and closing of hand.
- Timbre – Tilting of hand.

Once the users had become accustomed to controlling audio processes with hand gestures they were then instructed to make their own suggestions for mappings of the four aforementioned musical features. The interaction with the pre-mapped system could have biased the participants. The study was performed in a relaxed manner in which people controlled playful sounds as well as “live generated music (homophonic chorale)” (Berndt et al., 2016, p. 4) . 44 people participated in the study. This resulted in 281 suggestions of 115 gestures. Single hand gestures were preferred, with only 13.5% of the gestures being two handed. Each category received a number of popular gestures.

There were two popular gestures for tempo, a fanning motion with one hand and a circular movement of one hand that is held in front of the body. For the dynamics there were 65 suggestions for vertical hand movements where the gestures moved up to increase and down to decrease the volume level. This is a similar finding to Lech and Kostek (2013), as during their study up strokes controlled the intensity of any given parameter. For articulation the most popular gestures involved "grabbing/finger spreading and smooth/choppy movements to indicate articulation" (Berndt et al., 2016, p. 5). For the Dark/Soft timbral features participants performed a metaphorical wave gesture in which a horizontal hand with a palm pointing downwards imitates the motion of a wave. For bright/shrill timbral manipulation participants produced gestures that involved the spreading of fingers.

It is important to note that a number of the gestures that received the highest consensus were similar for a number of the different musical features. For example, the most popular gesture for increasing the tempo (open hand palm downwards), was also the most popular gesture for the crescendo and the third most popular gesture for the 'bright/shrill' timbre change. Therefore, Berndt et al. (2016) suggest altering the gestures slightly so that they were not exactly as they were proposed but with some minor changes in order to differentiate between the audio processes.

Although the majority of the musical features mentioned during this paper may not appear related to a DJs specific domain, the author believes that a number of these features could be feasibly transferred to a DJ specific domain. Clearly volume change gestures used in music production can be equivalently used in DJing. Maybe the gestures that work for spectral manipulation of content may work for the EQ in DJing, as they have some similarities.

This section has outlined the use of gestures for music production. All of these papers saw success in that they created an interface that were useable and intuitive (Lech & Kostek, 2013; Wakefield et al., 2017). Although there were a number a of negative factors, such as the gestures being uncomfortable (Lech & Kostek, 2013), and that the more complicated gestures were awkward to perform (Wakefield et al., 2017), these studies demonstrate that gestural control of audio parameters is not only feasible, but also affords greater expressivity in comparison to traditional mixing hardware. If these findings can be applied to the context of DJ performance, then gestures would afford better expressivity.

2.4.3 Live Performances

There has been a significant number of research projects that focus on creating sensors that translate movement (gestures) into a sonic output (Waisvisz, 1985; Bokowiec, 2011; Mitchell et al. 2012). These forms of technologies have been utilised since the emergence of tracking devices in the 1970's (Sturman & Zeltzer, 1994). The Theremin was the first mid-air gesturally controlled instrument which was released in 1928 by Leon Theremin (Theremin & Petrishev, 1996). This device consists of two metal antennas which create an electromagnetic field; when a hand moves within proximity of the antennas the movement is detected and musical notes are triggered.

As far as the author is aware the earliest instance of free hand gestural control of music is Michel Waisvisz's 'The Hands' (Waisvisz, 1985). The Hands is a "digital musical instrument centred on arm and hand movements for live performance" (Bellona, 2017, p. 1). Movement of The Hands controlled spatialization of the sound and various other effects. Michel Waisvisz developed The Hands over two decades. It is important to note that there were various iterations of The Hands, this analysis is from Movement 1 (Bellona, 2017).

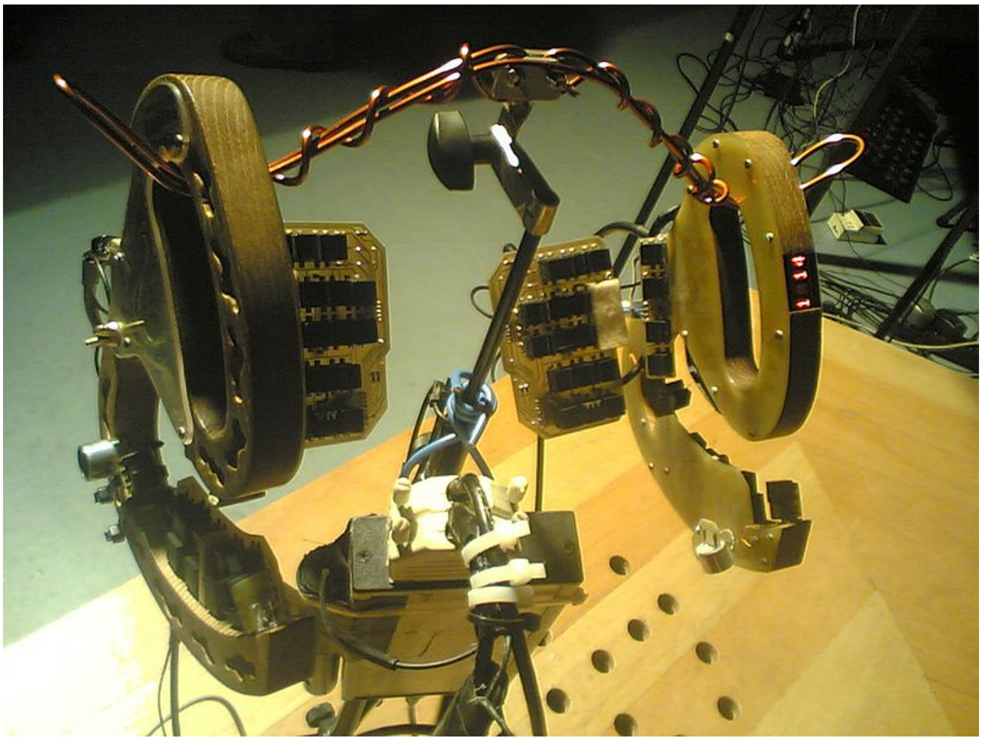


Figure 6. Michel Waisvisz's The Hands (Ivanovic, 2006)

The Hands consists of two devices that are worn over each hand (Figure 6). The left hand has two buttons near the thumb that control the step-up/down of MIDI programs. The MIDI programs control three Yamaha TX7 synthesizers which are each connected to different speakers: left, centre and right respectively. This results in the MIDI program having direct control over the location of the sound. The right-thumb button toggles 'scratch' mode (this has no connection to the DJ scratch). When scratch mode is active the active Note On event is copied resulting in the rapid generation of numerous notes. Bellona states that "since the controls of key-velocity and Note On messages are tied to the sonar sensor in 'scratch' mode, Waisvisz may only have had to move his hands apart to increase spectral richness of a sound" (Bellona, 2017, p. 4). The sonar sensors act as depth and distance perception which allow the accurate detection of the movement of each hand. The Hands utilised 10 different hand poses that toggled switches. Eight of the poses were used for active octave transpositions. The poses consist of: palm down, palm up, palm back, palm forward, palm left, palm right, palm down back right, palm down forward right, palm down forward left and palm down back left.

When 'scratch' mode is deactivated the distance between the two hands controls the dynamics. If the hands stay equidistant, then the dynamics stay fixed. However, if Waisvisz increases the distance then the volume increases and, depending on which synthesis algorithm is enabled, it can also add harmonic content. This enabled Waisvisz to easily create crescendos by slowing widening his arms. Waisvisz states that "The physical effort you make is what is perceived by listeners as the cause and manifestation of the musical tension of the work" (Krefeld & Waisvisz, 1990, p. 2), if one takes this statement literally and applies it to the current method of DJing, where little physical effort is applied, it becomes clear that a means of controlling sound through movement is required to enhance the overall audience experience.

Bokowiec (2011) created another piece of sensing hardware that has been developed through many years of iterative design (Wilson & Bromwich, 2001). Bokowiec (2011, p. 1) developed the Bodycoder System which is a:

> *"… sensor array designed to be worn on the body of a performer. It is a performance mechanism that enables a soloist to generate, affect, manipulate and control all aspects of a multimedia performance, comprising both audio and video material."*

Whereas Waisvisz's 'The Hands' were centred around sound creation through synthesizer(s), Bokowiec's Bodycoder System focuses on manipulation of the voice with a number of different spatialization techniques that are controlled through gesture. The voice manipulations are performed within Max/MSP through granular synthesis, which utilises Max's granulator object. The suit contains 12 switched inputs: four finger switches on the right hand that facilitate online and offline modes and 8 finger switches on the left hand that control certain Max/MSP patches. The main aspect of the suit that allows for gestural input are the bend sensors which are located on each elbow and wrist.

One of the main features of the suit is the spatialization. This works in two different modes, automated and performer controlled (through gestural interaction). The performer controlled mode operates in three different modes. When the first mode is activated, eight grain phases of a granulator are split into two channels which are then controlled by either of the wrist sensors. The second mode "… routes a mix of all grain phases to two rotational spatializers, the right wrist controlling a panning in an anticlockwise direction and the left wrist controlling a panning in a clockwise direction" (Bokweic, 2011, p. 3). When the third mode is active the right wrist sensor controls the pan speed of all eight of the granulator phases, which is set between 0 and 2500ms. It is stated throughout the paper that the creation of the interactions and the mappings was a strong collaborative process between the author and the performer. Therefore, the specific gestures were developed with this particular performer in mind. The gestures were not designed considering the production of the best action-sound mapping but were designed with affording the maximum amount of expressivity. This resulted in a gesture set that provided an expressive performance, but did not enhance the audience's understanding of the action-sound link and thus the understanding of how the performers actions relate to the sonic output.

Waisvisz left behind a legacy of performers who aimed to express themselves through body movements. 'Mimu gloves' (Mitchell & Heap, 2011; Mitchell et al., 2012) are similar to The

Hands in that they allow the performer to trigger audio samples and processes through hand movements. The Hands were tailored to Waisvisz needs, in contrast the Mimu gloves, although developed by a single performer (Imogen Heap) with her performance in mind, are now a commercially available product that can be mapped to any musical software that accepts MIDI/OSC messages. The Mimu gloves contain three hardware sensors: bend detection sensors on each finger, orientation sensors on the wrist and then a button on the index finger. The glove also contains a vibration motor on the reverse of the hand which is designed to provide haptic feedback. Figure 7 shows a live performer using a Mimu glove.

Figure 7. Kelly Snook performing with a MiMu glove at the Prix Ars Electronica Gala 2014 at Brucknerhaus (Ars Electronica, 2014).

During an interview Heap states that the gloves allow her to "… make music on the move, in the flow, and more humanly, more naturally…" (Dezeen, 2014, 00:32). Heap states that instead of increasing the level of a fader, she can simply raise her hand upwards, which is more expressive and exciting for not only the performer but audience members also (Dezeen, 2014). Heap then goes on to discuss her specific mapping. By raising and lowering her hand she can play notes of a scale. Due to the dynamic nature of the Mimu gloves mapping, the same gesture can control a number of different parameters (Mitchell & Heap, 2011; Mitchell et al., 2012). This raises the question: is there a universal gesture that is better suited to control a specific audio parameter? And if so how does one determine such gestures? For example, are more people likely to create a strong mental model for the action-sound relationship (and thus understand what the performer is doing with greater clarity) when a performer raises their hand and the volume raises. Or will there be a stronger understanding when the performer raises their hand and a low pass filter cut off increases. The fact that a solo musicians project that was designed to afford a new means of expression through gesture, has been developed into a commercially available product indicates the potential for gestural control over music.

Brown et al. (2018) performed an investigative study that focused on gesture design for the Mimu gloves. During this study four artists were interviewed regarding the design of the gestures that they use during their live performances. The overall consensus was that the participants used simple one-to-one and few-to-one mappings while adding ancillary gestures for aesthetic purposes. These simple mappings were easier to understand and thus created an optimum action-sound relationship.

These findings are contrary to traditional gesture design literature which stated that complex mappings were required to create performances that afforded maximum expressivity (Rovan et al., 1997; Hunt & Kirk, 2000; Momeni & Henry, 2006). Although the participants aided one an another in the development process the authors found that they created unique gesture sets that reflected “personal interpretations and aesthetics of music and movement” (Brown et al., 2018, p. 16). The gesture sets found in this paper support the common argument that metaphors should be used to aid the audience’s understanding of the performers actions. Each artist worked in an iterative design process in which the gesture sets were changing from performance to performance. This has very strong implications for this research. Due to this iterative process it begs the question: are there universal gestures that have strong action-sound links? Or do the performers require a more intimate relationship with the gesture set in which they build and develop on a personal level?

From the three performance devices discussed in this section a clear overarching theme has emerged that defines an iterative design process (Bokowiec, 2011; Bellona, 2017; Mitchell & Heap, 2011), in which the authors used performances as a of means testing the sensor device and gestures, and then improving the elements that they felt required updating in an incremental fashion. Consequently, these sensor devices were designed with a target user and thus specific performances in mind. This resulted in gesture sets that were tailored to the specific sensor device. Most examples of gesture systems that control music live are for individuals and not for the mass market and appear to be fundamentally personal. A major problem with this design process is the coupling of sensor device with gestures, and thus gestures with the specific sensor device. However, the aim of this research project is to create a gesture set that is universally understood. Therefore, the HCI discipline, which tends to focus on creating gestures to control commercially available products such as PCs, virtual reality (VR), smart home systems and TV, is drawn upon. This literature states that sensing devices such as the Leap Motion and Microsoft Kinect should not be utilised during the creation of gesture sets that are aimed to be used, and thus understood, by the masses. In the following section, existing research in creating gesturally controlled DJ systems is explored.

2.4.4 Gestures to control DJ performance

Several studies have considered redesigning the DJ interface. These interfaces range from: hardware interfaces for scratch DJs (Beamish et al., 2004; Slayden et al., 2005; Bryan & Wang, 2011), hardware interfaces for mix DJs (Villar et al., 2007, Dewey et al., 2018) to touchscreen interfaces (Hansen & Alonso, 2008; Lopes et al., 2010; Molina et al., 2011; Laursen et al., 2014). However, none of these interfaces consider the DJ-audience communication problem.

As far as the author is aware there have only been two attempts to create gesturally controlled DJ interfaces. Hayafuchi and Suzuki (2008) designed and developed a wearable glove that tracks hand orientation and finger movement for use in three musical applications: 'Air DJ', 'Air Conductor' and 'Wearable Music'. The Air DJ application allows users to control audio processing via hand gestures. The gestural input works by first striking a pose and then performing an action. Since it is not stated otherwise the gesture set was presumably designed by the authors. As stated previously, this style of gestural interaction is typical of gesture sets that are designed by a small team of designers (Norman, 2013). The paper lacks detail when discussing the DJ application stating that "When the user does some hand gestures or body motion, the computer translates them into musical control and produce audio sounds." (Hayafuchi & Suzuki, 2008, p. 2). Due to the lack of detail it is difficult to evaluate this system.

Sandor and Nakamura (2018) present a system that allows the user to scratch DJ with mid-air gestures. Similar to Hayafuchi and Suzuki (2008) this paper lacks detail therefore the system is difficult to evaluate. Both of the aforementioned papers fail to state the origin of the gestures and do not state the gesture design methodology. The studies also fail to evaluate the gesture sets.

2.5 The Development of Gesture Sets

The design of the gestures for the studies that have been discussed throughout this literature review have either been created by the authors or the origin of the gestures were not stated. In the latter case, it is fair to assume the gestures were either created by the authors or were pre-existing and have thus most likely been created by a team of designers. Norman (2013) states that these types of gestures are awkward and hard to learn and thus deter users from utilising new interfaces. If a DJ gesture set contained such gestures this would hinder the DJ further and may even degrade the perceived quality of a DJ performance. Whereas if an intuitive universally understood gesture set was designed, then DJ-audience communication could be improved; resulting in an overall better perceived performance.

Norman (2013) states that the awkward interactions that are caused by poorly designed gestures sets are down to not involving the end user in the design process. In his seminal book Norman goes into detail about how the everyday items that we interact with are often designed poorly and result in frustrating interactions which the users blames themselves for as opposed to the designers. He states that this can be applied to all means of interaction, including HCI. To combat the previously mentioned problems, Norman proposes a design methodology in which users are observed in their natural surroundings. By utilising such a design approach Norman claims that the frustrating interactions that we encounter on an almost daily basis can be eradicated. Norman (2010, p. 4) reiterated this point in an article discussing natural user-interfaces where he states that:

"Gestural systems are no different from any other form of interaction. They need to follow the basic rules of interaction design, which means well-defined modes of expression, a clear conceptual model of the way they interact with the system, their consequences, and means of navigating unintended consequences".

Another popular gesture design method is to simply allow users to define their own unique gestures. This method was detailed throughout Section 2.4.3 where a common theme emerged in which users defined their own gestures in conjunction with purpose built devices. This method has also seen some success in HCI (Morris et al., 2010).

In contrast to the previously mentioned methods, an elicitation approach could be adopted. By utilising a strict gesture elicitation approach gestures are more likely to fit a broader scope of people's mental models. The need for a methodology that produces a gesture set that fits the end-user's mental models is emphasised by Lopes et al. (2010, p. 2) when they state existing DJ interfaces that utilise gestural interactions force DJs to "learn a new set of gestures that may not easily match their mental models". During the next section the GES is detailed.

2.6 Gesture Elicitation and Identification Studies

The GES is a methodology that is used to acquire gesture sets from end-users in the hope of creating gestures that more accurately represent the needs of the subjects who are going to be utilising the system (Wobbrock et al., 2009). Henceforth, the term GES refers to Wobbrock et al.'s accepted Gesture Elicitation Study methodology.

Wobbrock et al. (2005) formalised a user-centred method that focuses on gesturally controlled interfaces. Four years later the first gesture set defined by utilising this methodology was created in the context of surface computing (Wobbrock et al., 2009). Surface computing uses a surface and tangible objects to control a computer, as opposed to standard input devices such as keyboard and mouse. Henceforth, the GES has become a popular tool for creating gesture sets that put the end-users' at the centre of the design process. This method has been applied to a number of areas which utilise different means of interactions such augmented reality (Piumsomboon et al., 2013), mobile devices (Ruiz et al., 2011), TV (Dim et al., 2016) and smart home systems (Lyons & Antle, 2018; Vogiatzidakis & Koutsabasis, 2019).

The basic principle of the GES involves showing the user the desired outcome that the gesture will map to (often referred to as a referent). During Wobbrock et al. (2009) study the referents are simple computer tasks such as zoom in, open, previous, next, accept etc... Then the user is instructed to perform the gesture they think will best execute the referent. Once the user has proposed the gesture they are then asked to perform the same gesture three times to act as a form of validation. If the user fails to produce the gesture three times, then they are asked to repeat the elicitation. Once the user has validated their gesture they're then instructed to answer two questions about ease of performance and gesture suitability. The think aloud protocol was used during the GES (Jääskeläinen, 2010). This technique entails the participants explaining their thought process while they propose each gesture. The subjective questions and the think-aloud protocol are seen as negative methods that could potentially remove the participant from their flow. Wobbrock et al. (2009) did not use sensing devices during the study as they are inherently restrictive which force users to learn a whole new means of interaction that detracts their attention away from the gesture set design.

This resulted in Wobbrock et al. using a prototype tablet which did not use Windows or Mac OS.

2.6.1 Grouping Gestures

Once the elicitation has taken place the identical gestures are then grouped together. The gestures that achieve the highest consensus constitute the gesture set. Therefore, consensus is used as the exclusive measure.

A debated topic within the HCI community is whether consensus should be utilised as the exclusive measure for the suitableness of gestures. A number of studies have stated that consensus is a suitable measure (Wobbrock et al., 2009; Morris et al., 2010; Dim et al., 2016). However, there has been some debate whether this is truly the case (Choi et al., 2012; Chen et al., 2018; Wu et al., 2019). Choi et al. (2012) performed a study that investigated whether end-users preferred the gestures that received the highest consensus during a GES. Firstly a GES was performed, then a month later the same participants selected their favourite gestures from the GES. Only 35% of the highest consensus gestures were chosen in the second study. This meant that 65% of the gestures that were selected in the second study were not the highest consensus gestures for the original GES which shows that consensus may not be the correct measure to use for selecting the suitableness of gestures. In contrast, in the original GES, Wobbrock et al. (2009) found that the gestures that were awarded higher subjective ratings tended to belong to the high consensus gestures. Dim et al. (2016) also found that the high consensus gestures were most preferred.

Several studies deviate from Wobbrock et al. (2009) rule of grouping identical gestures. During a number of GES, similar and identical gestures are grouped (Choi et al., 2012;

Piumsomboon et al., 2013; Chan et al., 2016). All of these studies place an emphasis on grouping the gestures which have the same thought process/mental model which aligns with the rational for this project.

2.6.2 Measuring Agreement between gestures

In an attempt to quantify GES results, Vatavu and Wobbrock (2015) proposed the agreement rating. Once the gestures have been grouped, the agreement rate can be used to create a quantification that can be compared to gesture groups from other GES.

$$AR(r) = \frac{|P|}{|P|-1} \sum_{P_i \epsilon P} \left(\frac{|P_i|}{|P|}\right)^2 - \frac{1}{|P|-1}$$

Equation 1.

Equation 1 displays the agreement rate (AR) formula, which ranges from 0 (no agreement) to 1 (total agreement). Where P is the number of all proposals for a referent r. P_i is a subset of the identical proposed gestures. For example, if a referent received the following: 6 gestures proposals, with the following number of people: 10, 7, 1, 1, 1, 1 for 21 participants, the agreement rate is computed below:

$$AR(Vol\ Up) = \frac{21}{21-1}\left[\left(\frac{10}{21}\right)^2 + \left(\frac{7}{21}\right)^2 + \left(\frac{1}{21}\right)^2 + \left(\frac{1}{21}\right)^2 + \left(\frac{1}{21}\right)^2 + \left(\frac{1}{21}\right)^2\right] - \frac{1}{21-1} = 0.314$$

Vatavu and Wobbrock (2015) proposed an agreement rate classification which enables researchers to discuss the agreement scores in qualitative terms (Table 2).

Table 2. Vatavu and Wobbrock's (2015) gesture agreement classifcation

$AR(r)$	Interpretation

$\leq .100$	$low\ agreement$
$.100 - .300$	$medium\ agreement$
$.300 - .500$	$high\ agreement$
$> .500$	$very\ high\ agreement$

2.6.3 Legacy Bias Mitigation

One of the largest potential drawbacks of performing GES is that the proposed gestures are often biased by the user's previous experiences. During HCI based GES, legacy biasing tends to produce gestures that are focused around mouse and keyboard interactions. Morris et al. (2014, p. 3) identify three reasons for legacy inspired gestures being elicited:

- A desire to transfer knowledge from past systems to new systems.
- A desire to reduce mental and physical exertion when interacting with new systems.
- A misunderstanding of the capabilities of input sensors.

Regardless of the reasoning behind such a phenomenon, legacy bias needs to be addressed in designing a DJ GES.

Numerous research projects that employ the GES methodology have reported a number of different problems caused by legacy biasing. Wobbrock et al. (2009) noted that despite providing a large touchscreen that lacked traditional UI elements, such as icons, the majority of participants still proposed mouse-like gestures. Even the participants acknowledged the legacy biasing with one stating "I'm falling back on the old things that I've learned" (Wobbrock et al., 2009, p. 8). During a multimodal gesture elicitation study Morris (2012) noted that one of the participants referred to their hand as a mouse.

On the other hand legacy bias inspired gestures can aid in guessability and learnability; especially in walk-up-and-use systems in public settings (Morris et al., 2014). DJ legacy controls consist of either faders or knobs, which are utilised for the control of multiple audio processors. The audience would be unable to differentiate between the audio processes if a DJ was to utilise a 'knob turn' or 'fader' gesture due to the use of the same hardware parameters for a number of different tasks; legacy bias inspired gestures are deemed negative for DJ focused GES.

Morris et al. (2014) propose several techniques to mitigate the effects of legacy biasing. Their production technique requires participants to propose multiple gestures for each referent. Morris et al. even suggest forcing participants to keep proposing gestures until they propose a gesture that no other participant had proposed. The production technique would increase test duration drastically.

The second technique that Morris et al. (2014) proposed requires the participants to undertake a 'priming' phase prior to the GES. The premise of this technique was to attempt to force the participant to think about the capabilities that new technologies afford and the possible interactions that this entails. Morris et al. define priming by suggesting that participants could be shown either a video or in an person demonstration of a variety of ways of interacting with the target sensing device. For example, the participants could be shown gestures that are designed by HCI professionals which tend to be more complex than the gestures proposed by the end-user; this could act as a way of inspiring new ideas that are completely removed from legacy inspired interactions. Finally, in order to fully immerse the participant in the creative process they could mimic the priming examples.

The final technique that Morris et al. (2014) proposed is partners. When utilising this technique the GES is hosted in groups of two or more people, as opposed to single participants. This works by participants ‘bouncing’ ideas of each other which could potentially aid in idea creation. Morris et al. (2014, p. 5) even propose a GES that is “based on popular games like Charades, in which a participant knows that their partner will need to be able to guess the meaning of their gesture”.

Morris et al. conducted a GES to explore the efficacy of employing a production and priming legacy bias mitigation technique. One group of participants were primed with a video that showed a number of different gestural interactions such as the gestures performed by a sports referee and an aircraft carrier signaller and friends waving to each other. A further subset of this group were primed kinaesthetically. This group were instructed to perform several physical actions such as toe touches, jumping jacks and pointing to different positions in the room. Then during the GES a production priming technique was utilised, the participants were prompted to propose as many gestures as possible. The GES took between 45-90 minutes for participants to produce gestures for 14 referents. This study found that the primed participants produced fewer legacy biased gestures than those who were not primed; however these results were not statistically significant. Furthermore, the resulting gesture sets still contained legacy inspired gestures. The participants who were kinaesthetically primed tended to produced gestures that involved more movement than the other groups but this trend also failed to reach statistical significance.

Cafaro et al. (2018) propose a specific priming technique. The priming consists of placing the participant in a ‘frame’; frames are scenarios which are made of “conceptual units that organize the basic experiences of our everyday life” (Cafaro et al., 2018, p. 2). The framing consists of three primary factors:

1. Visual priming - pictures or a video are displayed to participants.

2. Written task - participants are instructed to write about a certain scenario.
3. Embodied priming – participants are instructed to perform what they wrote in their written task.

Cafaro et al. (2018) then went on to test whether utilising a frame priming technique would improve the discoverability of a gesture set in an interactive museum instalment. Two groups of people were primed utilising different scenarios: a gym and a 'funhouse'. A third 'control' group performed the GES without performing a priming phase. The resultant gesture sets used different body parts depending on the priming scenario; the 'funhouse' scenario received over 50% more full body movements than the other groups. This example is directly relevant to this research project, as this project aims to create a full body gesture set. Therefore, the users should be framed within a scenario that utilises the whole body. Cafaro et al. (2018, p. 6) found that "participants recommended gestures and body movements that were interconnected with each other, because they were grounded on the same priming frame". During the framing the participants are placed into a scenario which results in a theoretical grounding that could potentially further immerse the user into the creative process; Morris et al. (2014) state that immersion can aid in the mitigation of legacy bias. Cafaro et al. (2018) also found that the number of gestures that were discovered by the museum attendees were significantly higher with the gesture sets that were elicited from the groups who had been primed using the framing technique.

2.6.4 Gesture Evaluation

The majority of GES do not evaluate their gesture sets. Wobbrock et al. first introduced the idea of validating elicited gestures by suggesting a reverse of the GES study where "Unlabeled video clips of the gestures can be shown to 20 new participants… to see if people can guess which gestures perform which commands" (2009, p. 9). Ali et al. (2019) then went on to formalise this methodology and coined it the identification study. Ali et al. presented an application (Crowdlicit) that allows researchers to host GES and identification studies via web browsers. The Crowdlicit application was then used to perform a GES and an identification study. Gestures and voice commands were accepted in the GES. During the identification study participants were asked to “imagine they were interacting with a TV-based web browser” and then they were displayed a text symbol or a voice command that represented the gesture and asked to “freely propose one referent in text form” (Ali et al., 2019, p. 7). Participants were able to identify all of the referents which was probably due to the abundance of voice commands which are easier to identify as they spell out their intended purpose.

There have also been a number of studies that evaluate full gesture sets, often by using a comparative method. As far as the author is aware there is no standardised method of evaluating gesture sets.

Morris et al. (2010) performed another gesture comparison study. This paper aimed to define whether end-users preferred expert designed gestures or user-elicited gestures. The authors designed a set of gestures to control the referents found in Wobbrock et al. (2009). A total of 44 gestures were defined; a one handed and a two handed gesture for each referent. Wobbrock et al. elicited gestures were used as comparison gestures. This resulted in a total 81 gestures.

The first step of the test was to display the name of the referent on a screen and then an audio command announced the referent name and gave a brief description. A video then displayed an actor performing one of the proposed gestures; the Wizard-of-Oz technique (Beringer,

2001) was utilised in the videos so that the gesture appeared to execute the referent. Then the participants were asked to perform the gesture seen in the video so they could physically judge which one they preferred. The participants were then presented with two separate questions based on goodness and ease of use. This process was then repeated for all of the proposed gestures.

Then, the participants were presented with ‘large thumbnails’ which represented each of the gestures for the respective referent. They were asked which of the gestures they felt best suited the current referent. If the user did not recognize the gesture from the thumbnail then they were able to access a video of the gesture being performed. This process was repeated for each referent. The test took anywhere in-between 60-90 minutes. The gestures that were proposed by both end-users and experts were most preferred. Furthermore, the gestures that were proposed by end-users alone were preferred to the gestures that were proposed by the experts. This means that end-user elicited gestures are preferred over expert design gestures but this study suggests that the most optimal method of creating gestures is to involve both end-users and experts in the gesture design process.

Nacenta et al. (2013) performed a study that evaluates whole gesture sets. During this study they compare gesture sets that were designed through different methods in the aim of defining the optimal gesture design methodology. Three different types of gesture sets were compared:

1. Stock gestures - a set was designed from gestures that were part of pre-existing devices.
2. Pre-designed gestures - a set was designed by the authors.

3. User-defined gestures, during the first phase the users were prompted to create their own gestures.

The testing was conducted in three phases: the first stage constituted the learning phase, which was performed following the GES methodology. For the pre-existing and the pre-designed sets the users were shown the referent and then shown the corresponding gesture. However, for the user-defined set, the participants were asked to propose their own gesture. 24 hours later the second session (the 'reinforcement' phase) was performed. During this phase the participants were shown a referent and asked to reproduce the respective gesture, the participant was then informed whether the gesture was correct. If the participant could not reproduce the corresponding gesture or if the gesture was incorrect, then the gesture was shown to the participant in the form of a video. The final stage was performed in the same fashion as the reinforcement phase, except that no video reminding participants of the right gesture was presented at the end of the trial, neither were the participants informed of the correctness of the gesture. The design method which was utilised to generate the gesture set which received the highest number of correct gestures was deemed the best gesture set design methodology when aiming to create a gesture set with optimal memorability.

17 out of 18 of the participants choose the user-defined as their favourite set (Nacenta et al., 2013). Therefore, the user-defined design method was the preferred design methodology for creating gesture sets. This study has shown that generating gestures with a high memorability produces suitable gestures to control primitive computer tasks. In contrast this thesis wishes to create gestures with strong action-sound relationships in order to better communicate how the DJ's actions relate to the sonic output.

2.6.5 Use of gesture elicitation within music

Two hundred and sixteen GES have been performed since its introduction in 2009 (Villarreal-Narvaez et al., 2020). Yet as far as the author is aware only two studies have performed gesture elicitation within music, and both of these studies did not follow the strict GES methodology (Caramiaux et al., 2011; Berndt et al., 2016). Berndt et al. (2016) was detailed in Section 2.4.2.

Caramiaux et al. performed a study to investigate gestural embodiment of environmental sounds. An elicitation study was performed to test the following hypothesis: "… causal sounds imply symbolic gestures and non-causal sounds induce morphologic gestures" (Caramiaux et al., 2011, p. 2). Symbolic gestures are gestures that mimic the action that has produced the sound, whereas morphologic gestures are gestures that follow (or trace) the temporal evolution of the perceived sound features.

Prior to performing the elicitation study a sound bank was created. Firstly, the causal sounds were downloaded from a database, then the authors synthetically created the non-causal sounds so that they sounded similar to their causal counterpart but the source of the sound could not be identified. 21 participants were then recruited. The participants were then played each non-causal sound and asked how well they could identify the source. Four of the sounds whose source could not be identified were chosen for the non-causal corpus. There causal counterparts were utilised to create the causal sound corpus.
The participants were then split into two separate groups. Eleven were assigned to causal sound corpus and the remaining ten were assigned to the non-causal (synthesised) corpus. Each participant was then presented with a sound from the respective corpus and told the following:

"You must perform a gesture associated to the sound you will listen to. Here "associated" means performing gestures that mimic the action producing the sound or that follow temporal evolution of the sound."
(Caramiaux et al., 2011, p. 3).

The examiners then showed the participant examples of sound producing and sound tracing gestures. The statement and example gestures could bias the participants' gesture proposals. The elicitation was performed in three stages:

- Training – the participant can listen to the sound as many times as they would like and thus any number of gestures can be performed for rehearsal purposes.
- Selecting – once the participant is happy they perform the respective gesture.
- Validating – the participant then has to perform the selected gesture three times in a row, in time with the sound.

After the elicitation was completed the candidates were interviewed, where they sat with an examiner and watched back the recordings of the gestures and were told to verbalise their action (Caramiaux et al., 2011), this is similar to the think-aloud protocol that Wobbrock et al. (2009) utilise in their GES. The aim of the interviews were to discover the participants mental models to see if there were any links between the participants' thought processes. The authors found that the gestures were arbitrary and did not have any relation to the sounds.

Neither of the two papers that have performed gesture elicitation in the music disciplinary have followed the strict GES methodology (Caramiaux et al., 2011; Berndt et al., 2016). Throughout this section the methods that potentially biased the participants of both of the studies have been outlined.

2.7 Conclusion

The first section of the literature review detailed three different methodologies that could be used to define common DJ techniques. An ethnographic method was selected in which third party footage is analysed using an inductive approach.

Then the communication that occurs during DJ performance was detailed; with a focus on DJ-audience communication. This section outlined three main techniques that DJs perform in order to communicate with their audience:

- Over exaggerated hand movements (which can be seen as a result of the degradation of control intimacy).
- Large universally understood dance movements/gestures.
- Angling equipment towards the audience for ease of visibility in order to display how specific actions are affecting the sonic output.

These factors are a testament to the DJ-audience communication struggle that traditional DJ hardware has created. The universal dance movements/gestures suggest that there is an existing language that both the DJ and audience members can equally understand.

Then gestures in music performance and production were analysed. The impact that the visual aspects of conductors can have on the perceived experience of an orchestral performance was highlighted; especially in relation to the conductor’s expression. These findings suggest that DJ performance would benefit from a tool that affords full body gestural control. Gestures to control music production showed the potential of gestural control of audio processes. This section was concluded by presenting the minimal work that has been performed in creating gesturally controlled DJ interfaces.

Penultimately, gesture design methodologies were detailed. Three design methods were outlined:

1. Small teams of experts designing gestures from their past experiences.
2. Users defined gestures.
3. The GES.

The first approach tends to create gestures that do not align with the end-user's mental models. The second approach creates gestures that are personal and are strongly linked with sensor devices. The aim of this project is to define gestures that are universally understood, therefore this personalised methodology is not adopted for this project. This project utilised the GES methodology for the following reasons: firstly the GES adopts a UCD approach which could aid in creating universally understood gestures, secondly the GES has successfully been used to create gesture sets for a number of different systems within HCI.

Finally the GES was outlined. In order to avoid biasing, sensing devices are not used during the GES. Consensus is used to measure the suitableness of gestures due to the number HCI studies that stated consensus is the most appropriate measure. Legacy biasing was outlined with a focus on three legacy bias mitigation methods. The most appropriate method for this project is to use a framed based priming, and to avoid the production and partner methods. The production method doubles the total duration of the test which is deemed unacceptable for this application domain. The partners method is not used due to the complicated nature of eliciting in groups. There is not a standard method of evaluating gestures however several gesture evaluation methodologies were outlined which were mainly used to test the efficacy of end-user gestures; adaptation of these methodologies are used during this project. Finally, the use of gesture elicitation within music was detailed. As far as the author is aware, no music based studies have previously used the strict GES methodology.

3 - Identification of DJ Techniques

3.1 Introduction

This section outlines the first experiment in this study which focussed on identifying common DJ techniques. The aim of this experiment was to identify the DJ techniques that would then be presented to users during the GES. As justified in Section 2.2 an ethnographic approach using videos of professional DJ performances was used during this experiment. This chapter begins by outlining the approach in detail before presenting and discussing the results.

3.2 Experiment Design

3.2.1 Video Selection Criteria

Once the medium for analysis was defined, a specification was detailed to refine the search for DJ performance videos:

- The DJ had to play either House or Techno music. There are two reasons for defining such a rule. Firstly, the author is experienced in DJing House and Techno music and it was felt that having an awareness of these two genres would aid the analysis. Secondly, these genres of music tend to be played in the majority of clubs in the western world, therefore the techniques that are used when mixing these genres are most representative of current DJ practices.
- The camera had to be constantly focused on the DJs hands and equipment throughout the video to make any parameter changes visible to the viewer and therefore codable.
- The footage had to be clear so that the orientation of the knobs could be easily discerned, this would allow the detection of the direction of the DJs movements.
- The video had to have 20000 views to prove the popularity of the DJ.

- To make the analysis more consistent with the real-world nightclub environment, the video had to be 40 minutes or longer, as DJ sets in nightclubs tend to last for more than 40 minutes. In addition, a DJ set that lasts any shorter than 40 minutes does not allow the DJ ample time to showcase their technique/style which places pressure on the DJ and can force the DJ to alter their approach to mixing.

YouTube was used for video retrieval due to its popularity and abundance of dance music orientated video channels. Over 100 videos were considered. The selection process resulted in ten videos been selected for analysis which are displayed in Table 3. The genres in the table contain some sub-genres which are derivatives of an amalgamation of two separate genres.

Table 3. The selection of DJ videos that are analysed

Video Name	DJ	Channel	Views (as of 22/12/2020)	Reference	Genre
Amelie Lens vinyl only home session	Amelie Lens	Amelie Lens	2800851	Lens (2017)	Techno
DJsounds Show 2016 - ANNA - special vinyl only set!	Anna	PioneerDjSounds	102403	Pioneer DJ Sounds (2016)	Techno
Cristoph - DJsounds Show	Cristoph	PioneerDjSounds	38005	Pioneer DJ Sounds (2018)	Progressive-house
Hot Since 82 Boiler Room x Warehouse Project DJ Set	Hot Since 82	Boiler Room	1504694	Boiler Room (2013a)	Tech-house
Huxley - DJsounds Show 2017	Huxley	PioneerDjSounds	28221	Pioneer DJ Sounds (2017)	Tech-house
Techno Mix 1st November 2019	Mark Jones	Mark Jones	27131	Jones (2019a)	Techno
#138 Tech House Mix November 23rd 2019	Mark Jones	Mark Jones	23649	Jones (2019b)	Tech-house
Paul Oakenfold - DJsounds Show 2014	Paul Oakenfold	PioneerDjSounds	197913	Pioneer DJ Sounds (2014a)	House
Skream Boiler Room London DJ Set	Skream	Boiler Room	1702596	Boiler Room (2013b)	House
DJ Sneak live 90s mix on vinyl - DJsounds Show 2014	DJ Sneak	PioneerDjSounds	500357	Pioneer DJ Sounds (2014b)	90’s House

Forty minutes of each video was analysed using inductive analysis; during this process the uses of nodes are referred to as references. Every time a DJ performed a task the respective node was either coded, or if the task had not been performed by any of the previous DJs, a new node was defined and then coded. If the DJ removed their hand from a parameter but kept returning to the parameter in quick succession with a clear intention of portraying the parameter changes as one cohesive adjustment, then this was coded as one movement. This grouping of small movements was performed because in these instances the DJ wanted the small movements to be heard as one continuous sonic manipulation. For example, when turning the wet/dry knob of the master effects up to increase the level of the master effect over the build, the DJ may keep removing their hand from the knob between small increases however this increase in effect level was intended to be perceived as one continuous increase. Contrastingly, small distinct movements that happen repeatedly and were intended to be perceived as distinct separate adjustments were coded as such e.g. when the DJ quickly increases then decreases the wet/dry of the master effects level with a reverb effect selected, to create a 'splash' effect.

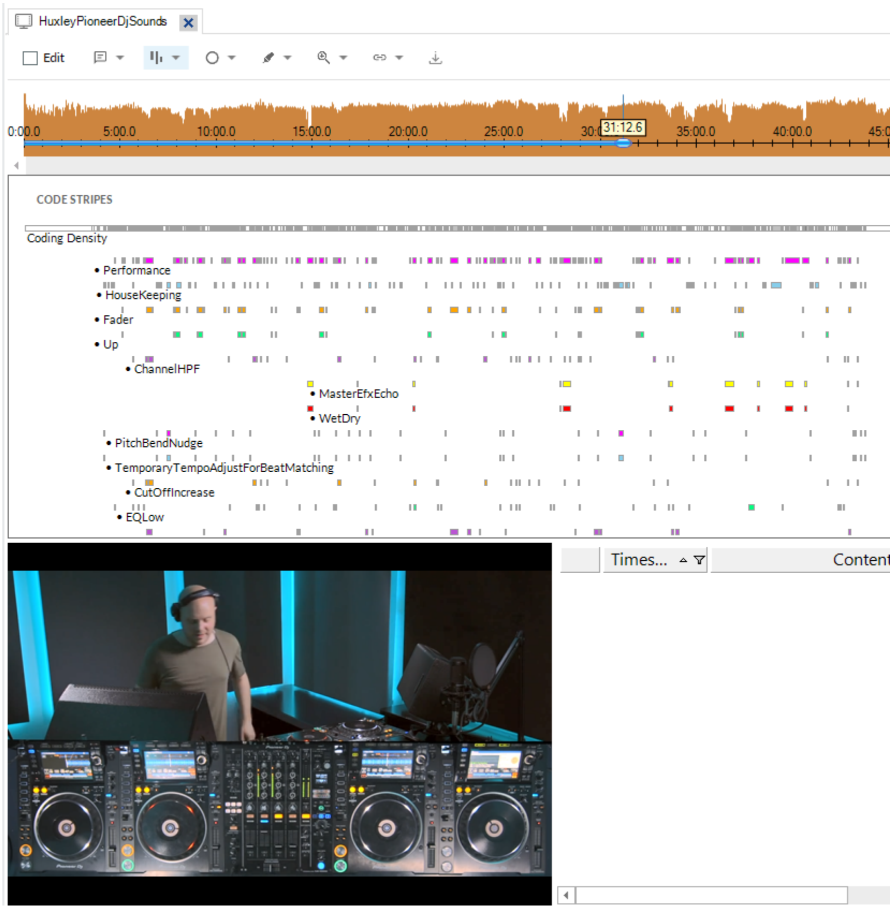


Figure 8. A screenshot of the DJ analysis Nvivo project. Waveform with superimposed timeline, hierarchical structured coding section and the DJ video under analysis from top to bottom respectively

Parameter changes that were made concurrently were also documented as a number of DJs manipulated two parameters concurrently to realise one sonic intention. Figure 8 shows the software environment used for documenting the DJ techniques.

3.3 Results

Four hundred minutes of video analysis resulted in 6827 references to 59 nodes which had a total duration of 240 minutes and 23 seconds. This means that on average the DJs were interacting with their equipment for 60% of their performance. The nodes were split into two categories housekeeping and performance. Housekeeping tasks did not effect the sonic output. Performance techniques effected the sonic output. The nodes were roughly made up of 50% housekeeping and 50% performance. During this analysis the housekeeping tasks were not taken into consideration as these actions did not contribute to the DJ-audience communication in relation to the sonic output. Once the data had been filtered, the list of performance based nodes were analysed in two different ways: number of occurrences and the total duration; this data is displayed in Table 4-8. Before detailing the selection of the DJ techniques the observations from the ten DJ performances are discussed.

Table 4. The filter and EQ performance DJ techniques

Group	Technique	Number of DJs	Number of Occurrences	Total Duration
ChannelHPF - Sub Total			**904**	**20:56**
ChannelHPF	Boost&Cut	2	28	00:40
ChannelHPF	CutOffDecrease	10	167	03:22
ChannelHPF	CutOffIncrease	10	321	10:41
ChannelHPF	Disable	10	176	02:33
ChannelHPF	Enable	10	212	03:40
ChannelLPF - Sub Total			**19**	**00:37**
ChannelLPF	Boost&Cut	1	3	00:09
ChannelLPF	CutOffDecrease	3	3	00:02
ChannelLPF	CutOffIncrease	3	6	00:20
ChannelLPF	Disable	3	3	00:02
ChannelLPF	Enable	3	4	00:04
EQHigh - Sub Total			**210**	**03:51**
EQHigh	Kill	2	2	00:07
EQHigh	Attenuate	8	120	01:58
EQHigh	Boost	8	88	01:47
EQLow - Sub Total			**745**	**23:56**
EQLow	Attenuate	10	292	12:24
EQLow	Boost	10	302	08:58
EQLow	Kill	10	134	02:17
EQLow	VaryingAttenuation	5	17	00:17
EQMid - Sub Total			**453**	**09:17**
EQMid	Attenuate	9	240	05:00
EQMid	Boost	9	198	04:04
EQMid	Kill	4	5	00:04
EQMid	VaryingAttenuation	4	10	00:09

Table 5. The fader performance DJ techniques

Group	Technique	Number of DJs	Number of Occurrences	Total Duration
Fader - Sub Total			**560**	**36:31**
Fader	BottomToTop	2	9	00:07
Fader	Chop	1	9	00:09
Fader	Down	10	167	16:16
Fader	Kill	10	56	01:00
Fader	Up	10	262	18:53
Fader	Varying	6	57	00:07

Table 6. The master effects performance DJ techniques

Group	Technique	Number of DJs	Number of Occurrences	Total Duration
MasterEfxBeatRepeat - Sub Total			**27**	**00:29**
MasterEfxBeatRepeat	WetDry	1	27	00:29
MasterEfxEcho - Sub Total			**242**	**12:50**
MasterEfxEcho	ChangeParameter	6	121	01:51
MasterEfxEcho	WetDry	9	121	10:59
MasterEfxFlanger - Sub Total			**8**	**00:08**
MasterEfxFlanger	ChangeParameter	1	3	00:02
MasterEfxFlanger	WetDry	1	5	00:05
MasterEfxControl - Sub Total			**166**	**01:37**
MasterEfxControl	MasterEfxOff	8	77	00:37
MasterEfxControl	MasterEfxOn	8	88	00:58
MasterEfxControl	MasterEfxTouchStrip	1	1	00:02
MasterEfxPhaser - Sub Total			**4**	**00:05**
MasterEfxPhaser	ChangeParameter	1	2	00:02
MasterEfxPhaser	WetDry	1	2	00:03
MasterEfxReverb - Sub Total			**124**	**07:23**
MasterEfxReverb	ChangeParameter	5	43	00:35
MasterEfxReverb	WetDry	6	81	06:49

Table 7. The loop and jog wheel/Vinyl performance DJ techniques

Group	Technique	Number of DJs	Number of Occurrences	Total Duration
Loop - Sub Total			**43**	**00:33**
Loop	Auto - 4 Beat	5	26	00:23
Loop	Auto - Double	1	3	00:03
Loop	Auto - Exit	4	9	00:05
Loop	Auto - Halve	1	1	00:01
Loop	In	2	2	00:01
Loop	Out	2	2	00:01
Jog Wheel/Vinyl - Sub Total			**8**	**00:05**
Jog Wheel/Vinyl	SlowScratch	1	1	00:01
Jog Wheel/Vinyl	Spinback	2	7	00:04

Table 8. The channel effects performance DJ techniques

Group	Technique	Number of DJs	Number of Occurrences	Total Duration
ChannelEfxBitChrusher - Sub Total			**4**	**00:05**
ChannelEfxBitChrusher	Disable	1	1	00:00
ChannelEfxBitChrusher	Enable	1	1	00:01
ChannelEfxBitChrusher	Increase Intensity	1	2	00:04
ChannelEfxEcho - Sub Total			**6**	**00:06**
ChannelEfxEcho	Decrease Intensity	1	1	00:01
ChannelEfxEcho	Disable	1	2	00:02
ChannelEfxEcho	Enable	1	2	00:02
ChannelEfxEcho	Increase Intensity	1	1	00:01
ChannelEfxNoise - Sub Total			**59**	**00:16**
ChannelEfxNoise	Decrease Intensity	2	3	00:04
ChannelEfxNoise	Disable	1	1	00:02
ChannelEfxNoise	Enable	1	2	00:02
ChannelEfxNoise	Increase Intensity	1	5	00:09
ChannelEfx Control - Sub Total			**24**	**01:08**
ChannelEfx Control	ChannelEfxWetDry	2	24	01:08

3.4 Discussion

There were several techniques that most of the DJs utilised throughout their sets. Firstly, attenuating the low frequencies when fading a track in. This is a standard mix DJ practice which allows for smoother transitions. This technique often includes the other EQ bands, however the DJs tended to focus on the low frequency band during this analysis; this is most likely due to the dominance of the low frequencies in House and Techno. On a number of occasions the DJs would increase the intensity of the HPF to around 50% in conjunction with the attenuation of the low frequencies, this emphasised the removal of the low frequencies. In contrast to the significant use of the HPF, the LPF was used minimally (number of occurrences for the HPF = 904 vs number of occurrences for the LPF = 19).

When transitioning between tracks the DJs mainly used the line faders (number of occurrences for the line fader = 506 vs number of occurrences for the crossfader = 10). This contrasts to the techniques employed by hip-hop/R&B DJs who use the crossfader heavily (Webber, 2018). This is due to the smooth transitions that are commonly found in Techno and House sets, as the line faders afford more precise control over the volume in comparison to the sharper volume curve of the crossfader .

The DJs mainly used echo and reverb as master effects (number of occurrences for echo + reverb = 364 vs number of occurrences for all of the other master effects = 40). This is most likely because these effects are characteristic of the genre. The most popular method of using these effects was during the build where the amount of effect is gradually increased until it peaks at the moment of the drop, then the effect is quickly removed. The reverb and echo effects were also used when transitioning tracks to help mask the upcoming track. Reverb was also used to create a ‘splash’ like effect that is often heard in dub reggae. This technique consists of the DJ engaging the HPF, then suddenly increasing the amount of reverb and then suddenly decrease the amount of reverb so that only a small time period, often a specific element such as the snare, of the track is allowed to feed the reverb.

The concurrently changing parameter changes were also taken into consideration when defining the common techniques. Four concurrently changing parameter changes (‘Echo + HPF - Increase wetness’, ‘Echo + HPF - Increase wetness then decrease delay time’, ‘Reverb + HPF - Increase wetness’ and ‘Reverb + HPF – Splash’) were commonly found across a number of DJ sets that were made up of parameters changes that received a high number of occurrences.

3.5 Technique Selection

When selecting the DJ techniques a higher precedence was placed on the number of occurrences as this provided an indication of how many times each was parameter was adjusted/used. This resulted in the selection of parameters that had a significant amount of occurrences, regardless of the duration. The selected DJ techniques can be found in Table 9. Throughout this project the techniques are often analysed in groups which can be found in the first column.

Table 9. The list common DJ techniques

Technique Group	**Technique**	**Criteria**
Filter	HPF - Sweep up then down	Number of occurrences
Filter	HPF - Sweep up	Number of occurrences
Filter	HPF - Sweep down	Number of occurrences
Filter	LPF - Sweep up then down	Total duration
EQ	EQLow - Kill	Number of occurrences
EQ	EQLow - Attenuate	Number of occurrences
EQ	EQLow - Boost	Number of occurrences
EQ	EQLow - Vary attenuation	Number of occurrences
EQ	EQMid - Kill	Number of occurrences
EQ	EQMid - Attenuate	Number of occurrences
EQ	EQMid - Boost	Number of occurrences
EQ	EQHigh - Kill	Number of occurrences
EQ	EQHigh - Attenuate	Number of occurrences
EQ	EQ High - Boost	Number of occurrences
Fader Slow	Fader - Fade in	Number of occurrences
Fader Slow	Fader - Fade out	Number of occurrences
Fader Slow	Fader - Volume up	Number of occurrences
Fader Slow	Fader - Volume down	Number of occurrences
Fader Fast	Fader - Kill	Number of occurrences
Fader Fast	Fader – Vary	Number of occurrences
Effect	Echo + HPF - Increase wetness	Number of occurrences
Effect	Echo + HPF - Increase wetness then decrease delay time	Number of occurrences
Effect	Reverb + HPF - Increase wetness	Number of occurrences
Effect	Reverb + HPF - Splash	Number of occurrences
Loop	Loop - Engage 4 beat	Number of occurrences
Loop	Loop - Engage 4 beat then decrease until $1/64^{th}$ of beat	Number of occurrences
N/A	Spinback	Exception
N/A	Pause Play	Exception

Once the techniques with the highest occurrences had been selected, the techniques with high total duration but low occurrences were selected. The only technique that was selected using this method was ‘Channel LPF – Sweep Up then Down’.

Having selected the common DJ techniques based on number of occurrences and total duration, ‘Spinback’ and ‘Play/Pause’ were added to the referents. The ‘Spinback’ technique only occurs eight times in all of the DJ performances; due to its distinctive sonic nature it is used sparingly. However, this technique is idiosyncratic to DJing. ‘Play/pause’ fell into the housekeeping category as DJs only ever played and paused tracks that were muted whilst the track was cued in their headphones; having no effect on the sonic output. Therefore, by default the play/pause parameter change was not included in the list. However, playing and pausing tracks are an important part of a DJ performance. Therefore ‘Play/Pause’ and ‘Spinback’ were included in the referents.

3.6 Conclusion

To conclude, an ethnographic approach has been utilised to identify common DJ techniques. This method involved the analysis of third-party footage using an inductive analysis approach. These techniques will be used in the GES which is detailed in the following section.

4 - Gesture Elicitation Study

4.1 Introduction

This chapter details the main experiment undertaken in this project. Once the identification of DJ techniques had been completed, and the resulting techniques had been utilised to create the referent audio samples, the GES could take place. The process of creating the audio referents is detailed in the experiment design section. The aim of this experiment was to elicit universally understood gestures to improve DJ-audience communication.

During the GES a strict process was followed in an attempt to prevent biasing. The names of the referents were withheld from the participants as the non-technical participants would not have the requisite knowledge to understand. This also forced participants to listen to the processing, which may encourage participants to define their own mental models. In order to avoid further biasing the use of sensor devices was also avoided. Legacy bias mitigation techniques were utilised in order to avoid participants' past experiences negatively influencing their gesture proposals. A strict context of use rational was followed in an attempt to make the participants feel like they were in the environment in which the gestures would be utilised (Maguire, 2001; Norman, 2013).

Firstly the methodology is outlined. Then the experimental design is detailed including the preparation of the audio stimuli (referents) and test environment. Then the conduct of the experiment is outlined and the results are discussed. Finally the section is concluded with a discussion of the elicited gestures.

4.2 Experiment design

4.2.1 Adaptations for this Application Domain

The Wobbrock et al. (2009) GES was performed in a lab setting. Subsequently the majority of GES also appear to have taken place in similar settings (Ruiz et al., 2011; Piumsomboon, 2013, Nacenta, 2013). However, Norman (2013) states that when analysing users they should be observed in their natural environment, thus the author believes the GES should be performed in the context that the gestures are intended to be used in. Following the same rationale as Norman (2013) the GES detailed in this chapter will follow the 'context of use' paradigm (Maguire, 2001); resulting in the GES taking place in a live music/club environment.

Wobbrock et al. (2009) advise using the 'think aloud protocol' during a GES. For the GES performed in this research the think aloud protocol was omitted for three reasons. Firstly, such a process could remove the participants from their flow. Secondly, due to the fact that music is constantly playing recording such data would be impractical. Finally, the think aloud protocol is arguably analogous to asking someone to explain the mental process whilst running up a flight of stairs. Humans are not consciously aware of how they perform primitive movements. They simply will a movement to happen and the subconscious mind executes the movement. Jääskeläinen (2010, p. 1) supports this argument stating that "Only information that is actively processed in working memory can be verbalized, which means that unconscious processing is inaccessible.". For these reasons the think aloud protocol was deemed unsuitable for this experiment.

Wobbrock et al. (2009) also asked test subjects to subjectively rate their gestures based on 'goodness' and 'ease' using a seven-point Likert scale however these subjective ratings were used for evaluation purposes and were not used when selecting the gestures. Therefore, for the application of a GES for a DJ gesture set, the subjective ratings were deemed unnecessary.

4.2.2 Preparation

There were three major aspects that involved preparation: the referents, the testing environment and the priming techniques.

4.2.2.1 Referents

Audio samples were generated that acted as referents during the GES. The audio samples (referents) would be presented to the test subjects who would then produce a gesture which they felt best executed the referent. Pre-existing commercial tracks were used for the creation of the referents. Multiple tracks were selected to avoid certain sonic elements from the individual tracks influencing the participants' gestures. Using multiple tracks also allowed different (sub-) genres to be represented in the study.

A list of requirements was created to help to refine the search for the tracks:

- The tracks must be professionally produced and mastered so that a DJ could credibly play them.
- The tracks must contain a broad distribution of frequencies so that subtle changes in all three of the frequency bands (low, mid and high) could be clearly discerned.
- The tracks must be produced in the genre of House or Techno (or one of the sub genres e.g. Tech-House, Minimal-House, Deep-House etc…) to mirror the tracks that were used in the DJ sets that were analysed during the identification of common DJ

techniques. These style of tracks can be found in the majority of clubs throughout the Western World, further helping to ensure that the context of use is realistic.

Table 10. The tracks that were used to create the audio referents

Name	Artist	Genre	Reference
Washed Away (Original Mix)	DJOKO	House	DJOKO (2020)
Ayahuasca (Original Mix)	Pavel Petrov	Tech-House	Petrov (2016)
Recreate (Slam Rework)	Hertz	Techno	Hertz (2019)

The tracks that were used to create the referents are shown in Table 10. Once the tracks had been selected the location of the referents in the track and the duration of the referents were defined. This was performed in an attempt to mimic the techniques that were identified in the previous experiment. A list of the locations and the durations of the referents in the track can be found in Table 11. The referents contained the same amount of dry and processed signal, so that a set amount of unprocessed signal was followed by the same amount of processed signal. For example, 'HPF – Sweep up then Down' contained two bars of unprocessed signal then the filter swept for two bars. This was performed so that the participants could compare and contrast the processed and unprocessed signals to be fully aware of the processing.

A Pioneer XDJ 700 (Pioneer, n.d.a) and a DJM 800 (Pioneer, n.d.b) were utilised to create the referents. The referents were recorded in mono as the majority of nightclub sound systems run in mono to be consistent with the 'context of use'. The referents were then imported into Adobe Audition and loudness normalised to -14dBLUFS. Each of the tracks were used to create the referents which resulted in three sets of referents.

Table 11. The audio referents with the location of the referent in the track and the duration of the referent

Technique Group	**Technique**	**Technique Description**	**Location in track**	**Processed duration**
Filter	HPF - Sweep up then down	Sweep the HPF up to the max then down to min	Main section (after drop)	2 bars
Filter	HPF - Sweep up	Sweep the HPF up to the max then disable	Main section (after drop)	2 bars
Filter	HPF - Sweep down	Sweep the HPF from max down to mix then disable	Main section (after drop)	2 bars
Filter	LPF - Sweep up then down	Sweep the LPF up to max then down to min	Main section (after drop)	2 bars
EQ	EQLow - Kill	Suddenly attenuate EQ low from 0dB to -26dB	Main section (after drop)	1 bar
EQ	EQLow - Attenuate	Gradually attenuate the EQ low from 0dB to -20.8dB	Main section (after drop)	2 bars
EQ	EQLow - Boost	Gradually boost the EQ low from -20.8dB to 0dB	Main section (after drop)	2 bars
EQ	EQLow - Vary attenuation	Vary the attenuation of the EQ low from -26dB to 0dB in rhythmical intervals	Main section (after drop)	2 bars
EQ	EQMid - Kill	Suddenly attenuate EQ mid from 0dB to -26dB	Main section (after drop)	1 bar
EQ	EQMid - Attenuate	Gradually attenuate the EQ mid from 0dB to -20.8dB	Main section (after drop)	2 bars
EQ	EQMid - Boost	Gradually boost the EQ mid from -20.8dB to 0dB	Main section (after drop)	2 bars
EQ	EQHigh - Kill	Suddenly attenuate EQ high from 0dB to -26dB	Main section (after drop)	1 bar
EQ	EQHigh - Attenuate	Gradually attenuate the EQ high from 0dB to -20.8dB	Main section (after drop)	2 bars
EQ	EQ High - Boost	Gradually boost the EQ high from -20.8dB to 0dB	Main section (after drop)	2 bars
Fader Slow	Fader - Fade in	Slowly boost volume line fader from 0 to 10	Intro	12 bars
Fader Slow	Fader - Fade out	Slowly attenuate volume line fader from 10 to 0	Outro	12 bars
Fader Slow	Fader - Volume up	Boost volume line fader from 3 to 10	Outro	8 bars
Fader Slow	Fader - Volume down	Attenuate volume line fader from 10 to 3	Main section (after drop)	8 bars
Fader Fast	Fader - Kill	Suddenly attenuate volume line fader from 10 to 0 on the 7th beat for one beat	Main section (after drop)	1 beat
Fader Fast	Fader – Vary	Alternate the fader from 10 to 0 in rhythmical intervals	Main section (after drop)	2 bars

Effect	Echo + HPF - Increase wetness	Activate HPF at 1/3 intensity then increase echo wetness from min to max	Build	8 bars
Effect	Echo + HPF - Increase wetness then decrease delay time	Activate HPF at 1/3 intensity then increase echo wetness from min to max while lower delay length at rhythmical increments until at min	Build	8 bars
Effect	Reverb + HPF - Increase wetness	Activate HPF at 1/3 intensity then increase reverb wetness from min to 2/3	Build	8 bars
Effect	Reverb + HPF - Splash	Activate HPF at 1/3 then increase wetness of reverb for one beat	Intro	1 beat
Loop	Loop - Engage 4 beat	Set a 4 beat (1 bar) loop	Main section (after drop)	6 bars
Loop	Loop - Engage 4 beat then decrease until 1/64th of beat	Set 4 beat (1 bar) loop then decrease loop length at rhythmical intervals until loop length is 1/64th of a beat	Main section (after drop)	6 bars
N/A	Spinback	Spinback for 1 beat then grab edge of platter to stop (while in ‘slip’ mode*)	Main section (after drop)	1 beat
N/A	Pause Play	Pause for 1 beat then play again (while in ‘slip’ and ‘vinyl break mode, vinyl break mode emulates the stopping of a record on a turntable)	Main section (after drop)	1 beat

* Slip mode allows the respective track to keep playing in the background while performing either a pause or a scratch/Spinback on the platter, this means that when one plays the track or releases the jog wheel the track will still be in the same position as if it had kept playing (Gizmo, 2009).

4.2.2.2 Test environment

4.2.2.2.1 Location Selection

The Live Music Production (LMP) room based at the Bluerooms Studios on the University of Huddersfield campus was chosen to host the experiment (Blue Rooms Online, 2016). The room is intended to be used to practice live sound engineering and has the capabilities to host a full band. The room contains a 7kW professional PA system which consists of two Logic Systems CS1290SB PA speakers, two Logic Systems CS1290B PA speakers, and two Logic Systems CS1296 PA speakers. The room also contains two lighting rigs. The room was designed to imitate a live venue scenario in order to align with the context of use.

4.2.2.2.2 Room preparation

Tests were hosted over a two week period during February 2020. The room was prepared in the following manner: The LED stage lighting were set to a static lighting pattern and adjusted until they looked similar to the lighting in a club and afforded visibility of the participant. The sound system was set to 80dB using a sound pressure level meter. The listening ‘sweet spot’ was found using trial and error: a track was played on the sound system and different listening positions were trialled and compared so that the test subject would receive the best possible listening experience. A cross was placed on the floor with tape to signify the ‘sweet spot’ and to indicate where the participant should stand.

A computer with Logic Pro X running was placed to the side of the room and this was where the author sat throughout the experiment (Figure 9). Finally a camera was placed directly in front of the cross to record the participants movements. Figure 9-10 show the LMP room prepared in the test state.

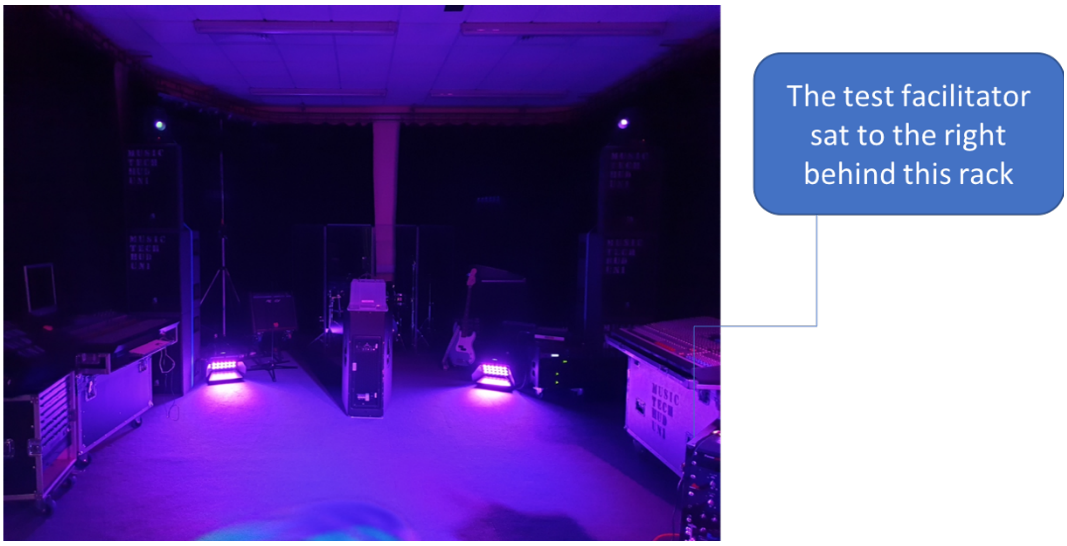


Figure 9. The GES test environment from the point of view of the test subject

Figure 10. The GES test environment

4.2.2.2.3 Priming preparation

Firstly, a priming video compilation was developed that displayed a multitude of gestures from non-DJ real life scenarios. The video contained snippets of public figures who are flamboyant gesticulators (including Benito Mussolini and Boris Johnson), avant-garde dancers, people signing in sign language, and people giving speeches with extensive use of hand gestures.

A frame-based 'Embodied priming' scenario was then prepared (Cafaro et al., 2018). This framing would take please after the priming video had been viewed. In attempt to immerse each participant in the club environment a story was written that detailed a person's visit to a nightclub. A secondary aim of the exercise was to make the participant comfortable with performing gestures in the authors presence. The story would be read to each participant who would be instructed to re-enact the story as it was being told. The story can be found below:

Matt goes to the club

Matt orders a taxi

The taxi arrives and matt gets in

The taxi arrives at the club, Matt gets out and greets some friends

Matt pays in to the club

He walks to his favourite area and dances

Matt walks to the bar and buys a drink

Matt dances some more

Matt decides to leave the club, he walks out and hales a taxi

The taxi takes Matt home, he gets out and goes straight to bed

4.3 Testing

4.3.1 Test 1 (live music environment)

21 test participants undertook the GES. 17 of the participants were male and 4 were female. The mean participant age was 25 years old. 13 of the participants listened to dance music. 8 of the participants were experienced DJs. 13 of the participants had experience with music technology (they either mixed and/or produced their own music). The author acted as the test facilitator. The first six participants were assigned the referents from the tracks by DJOKO (2020), the following six were assigned the referents from the track by Petrov (2016), the following six were assigned the white noise and then the remaining three participants were assigned the referents from the track by Hertz (2019). Upon entering the room the participants were read the following paragraph:

> *The gestures in the following video are just an example. Once you have watched the video I am going to read you a story line by line, once I have finished each line I would like you to perform the actions that you feel best represents the respective line of the story, if you feel uncomfortable I shall join in with you.*

The participant was then played a gesture compilation video (see https://selene.hud.ac.uk/gesturevideo/OtherVideos/GesturesCompilation.mp4). After the video had been viewed, the participant was then read a story line by line by the facilitator. After each line the participant was prompted to perform the corresponding action. Then the participant was read the following instructions:

> *"I am going to play you some audio examples which have been processed by professional DJ equipment, I would like you to make a gesture for the process, you can spend as long as you want on each gesture. Once you have found the gesture that you feel best suits the process, I would you like you to perform the gesture three times, exactly as you did the first time. The*

audio is processed rhythmically, in either periods of: 1, 2, 4, 8, 12, or 24 bars. If you do not understand when the process is occurring inform me of your difficulties and I can either signal when the processing is happening or if you are struggling further I can explain what the processing is. The gesture can start before and carry on after the processing. Before we begin I am going to play you an example of a track that's been processed and then I will play the three small snippets of the track that is being used throughout the test".

Before the test began two points were reiterated by the facilitator. Firstly, the participant was advised to choose the gesture that first came to their mind, the one that feels most natural and innate. Secondly, the participant was informed that there are no right or wrong gestures, all gestures would be taken into consideration during this study. The facilitator then positioned himself behind a computer screen to the side of the participant; the facilitator was mostly hidden but if the participant required help then the facilitator was in a position to offer his assistance. The test then began and the referents were played from a pre-randomised Logic Pro X project.

Each participant was given time to listen to the referent and then propose their gesture. If the participant wasn't aware of when the processing was happening, the facilitator held up a card with the words 'processing' written on. If the participant still wasn't aware of the processing, the referent was then explained. The referent was left on loop to play constantly as music in a club would be constantly playing, this meant that the unprocessed and processed sound would alternate by the number of bars defined earlier in this chapter (Table 11). For the longer referents (the *Fader - Fade in* and *Fade – Fade out*) after the participant had been played the unprocessed and processed sounds once, the participants were given the option to be played only the processed sound in an attempt to shorten the overall test duration. This was possible as it is clear what processing is occurring during these referents which removes the need for comparison. Once the participant had proposed the gesture, they were then instructed to perform the exact same gesture three times. On a number of occasions the participant was unable to perform the exact same gesture three times. During this situation the participant was instructed to restart the elicitation process for the current referent. This method was then followed for all of the referents.

4.3.1.1 Post-test survey

Once the test was completed each participant was asked a series of post-test questions (see Appendix A). The questions were designed to ascertain the participants involvement with dance music.

4.3.1.2 Multi Transition Referents

Initially two *Multi transition* referents were included in the DJ techniques. The first *Multi transition* referent consisted of a *Fader – Fade in* and a *EQLow – boost* and the second *Multi transition* referent consisted of a *Fader – Fade out* and a *EQLow – attenuate*. After completing the GES for the first six participants the ‘Multi transitions’ were removed from the experiment as the elicited gestures were mimicking the sub components of the referents. For example, the participants were proposing the same gestures that they proposed for *Fader – Fade in* and *EQLow – boost*. The *Multi transition* referents also caused confusion among a number of the participants which increased the overall test duration which was deemed unacceptable.

4.3.1.3 Alternate Referents

After performing the club environment based testing one major finding was observed: participants were eliciting gestures in relation to features of the music instead of the audio processing. For example, participants proposed gestures that imitated a drummer performing a snare roll which was a transient feature of the build in the music. In an attempt to isolate the audio processing from the music, a group of referents were made by applying the audio processes to white noise.

The white noise referents were created using Logic Pro X. Listening to white noise can be fatiguing. Therefore, in order to minimise the participants exposure to white noise, the number and duration of referents was reduced. The white noise referents can be found in Table 12. Constant white noise was utilised for the non-time-based effects, and pulsed white noise was used for the time-based effects so that the effects could be heard.

Table 12. The white noise based referents

Technique Group	Technique	Technique Description	White Noise Type	Processed Duration (Seconds)
Filter	HPF - sweep up then down	Sweep the HPF up to the max then down to min	Constant	7s
Filter	HPF - sweep up	Sweep the HPF up to the max then disable	Constant	7s
Filter	HPF - sweep down	Sweep the HPF from max down to mix then disable	Constant	7s
Filter	LPF - sweep up then down	Sweep the LPF up to max then down to min	Constant	7s
EQ	EQLow - kill	Suddenly attenuate EQ low from 0dB to -26dB	Constant	2s
EQ	EQLow - vary attenuation	Vary the attenuation of the EQ low from -26dB to 0dB in rhythmical intervals	Constant	7s
EQ	EQMid - kill	Suddenly attenuate EQ mid from 0dB to -26dB	Constant	2s
EQ	EQHigh - kill	Suddenly attenuate EQ high from 0dB to -26dB	Constant	2s
Fader Slow	Fader - fade in	Slowly boost volume line fader from 0 to 10	Constant	7.5s
Fader Slow	Fader - fade out	Slowly attenuate volume line fader from 10 to 0	Constant	7.5s
Fader Slow	Fader - volume up	Boost volume line fader from 3 to 10	Constant	3.75s
Fader Slow	Fader - volume down	Attenuate volume line fader from 10 to 3	Constant	3.75s
Fader Fast	Fader - kill	Suddenly attenuate volume line fader from 10 to 0 on the 7^{th} beat for one beat	Constant	.5s
Fader Fast	Fader – vary	Alternate the fader from 10 to 0 in rhythmical intervals	Constant	7.5s
Effect	Echo + HPF - increase wetness	Activate HPF at 1/3 intensity then increase echo wetness from min to max	Pulsed	15s
Effect	Echo + HPF -increase wetness then decrease delay time	Activate HPF at 1/3 intensity then increase echo wetness from min to max while lower delay length at rhythmical increments until at min	Pulsed	15s
Effect	Reverb + HPF - increase wetness	Activate HPF at 1/3 intensity then increase reverb wetness from min to 2/3	Pulsed	15s
Effect	Reverb + HPF - splash	Activate HPF at 1/3 then increase wetness of reverb for one beat	Pulsed	15s
Loop	Loop - engage 4 beat then decrease until $1/64^{th}$ of beat	Set 4 beat (1 bar) loop then decrease loop length at rhythmical intervals until loop length is $1/64^{th}$ of a beat	Pulsed	15s

4.4 Online Testing

After completing the GES for nine participants, the UK COVID-19 lockdown was announced; forcing the closure of the university campus and stopping further participants from performing the GES. Therefore, in order to complete the experiment, the methodology was modified.

4.4.1 Preparation

4.4.1.1 Application development

Due to the complicated nature of the GES it was clear that the facilitator would need to be present during the online testing to answer any questions and assist and direct the participant when necessary. This was performed via a video calling service (Skype, Zoom, FaceTime). An online web application was built to give participants access to the referents. The home page of the application contained the 'priming' gesture compilation video (Figure 11). Two buttons were displayed under the gesture compilation video. The first button redirects the test subject to the web player that contains the referents. The second button plays an example of a referent. Due to the fact that the participants would be performing the experiment in their lockdown location, the 'context of use' paradigm was no longer an option. Therefore, the framing priming activity, in which the participants were read a story and then they acted the story out, was omitted. Figure 11 displays the first page of the web application and Figure 12 displays the web player.

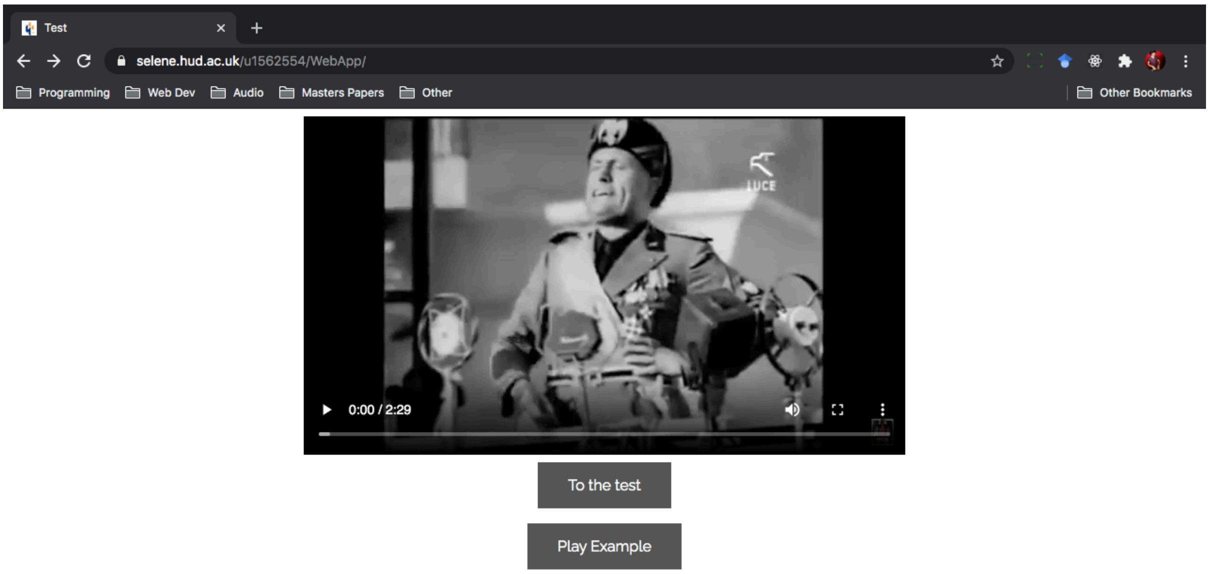


Figure 11. The home page of the website. The gesture compilation video is displayed at the top, then directly below the buttons redirect the user to the web player and play an example respectively

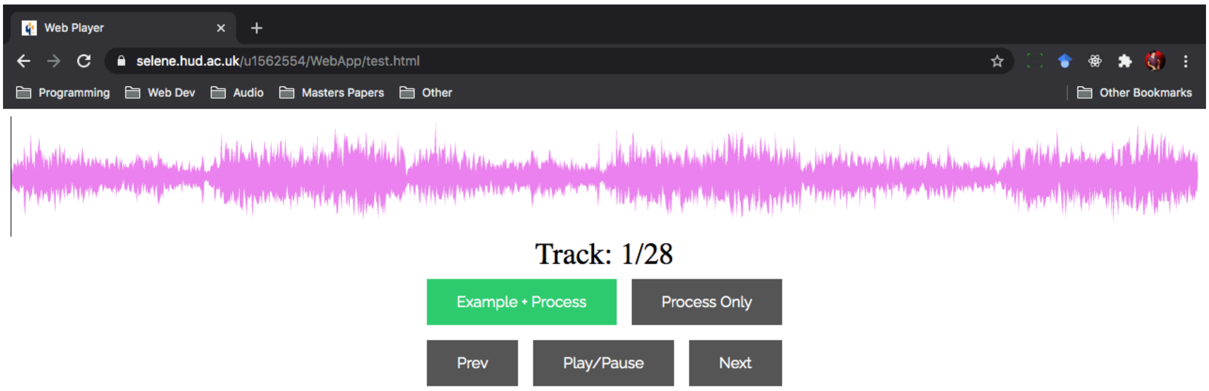


Figure 12. The web player with rudimentary playback controls. The Process only button removes the unprocessed signal

4.4.2 Test 2 (online)

Upon calling the participant via their preferred service (e.g. Zoom, Teams, WhatsApp, Skype), the facilitator introduced himself and the participant was prompted to visit the web application. The facilitator then read the same instructions as for the simulated club environment. The participant was then asked to watch the gesture compilation video which is found on the homepage of the web application. After watching the video they were asked to listen to an example of a referent by clicking the button underneath the video (Figure 11). Once the participant fully understood what processing occurred during the example they were then asked to proceed onto the web player. From here the interface was explained and the testing began. The session was recorded using the QuickTime screen recording function (Apple, n.d.). The elicitation then proceeded in the same fashion as described in Section 4.3.1, with the small exception of the participant being in control of the playback of the referents. After completing the GES the same post-test questionnaire was completed.

4.5 Preparation of Videos for Analysis

In order to perform the grouping of the gestures a means of comparing videos for each referent was required. Video compilations were edited that displayed all the participants for each referent group on one screen (Figure 13).

Figure 13. The first six GES participants who elicited gestures for the DJOKO (2020) based referents

4.6 Gesture Grouping

To begin the gestures must be grouped. The grouping process consisted of grouping together similar gestures elicited from different test subjects for each referent. The aim of the grouping is to create groups of gestures from the gestures that receive proposals from more than one participants. Furthermore, the group that contains the most number of gestures is deemed the best gesture. During the original GES the grouping of gestures is performed by comparing each gesture for a given referent, the identical gestures are placed into groups (Wobbrock, 2009). Due to the wide variety of gestures that were proposed throughout the GES, this experiment followed a similar approach as Choi et al. (2012), Piumsomboon et al. (2013) and Chan et al. (2016) who group similar and identical gestures.

Initially a text analysis approach was adopted where gestures were broken down into sub components performed by different body parts. However, it quickly became clear that such an approach made the gestures harder to compare due to the large amount of data that the text analysis produced. Then a more simple text analysis was performed in which simple overviews were written for each gesture. However, this method also produced a large amount of data which made the grouping process more complicated.

Therefore, a simple visual analysis process was followed. The author and both supervisors met over a two day period to perform a group consensus session. All gestures were visually compared to each other and grouped based at the discretion of the three academics. Through discussion in the sessions a consensus between the three researchers was reached. Firstly, gestures that were very similar were grouped. These gestures were awarded a point for each gesture. For example, for the *Fader - Fade in* referent six of the participants raised a single hand upwards, all differing in the range of the movements, some starting at waist height and others starting at knee height. The range of the movement was disregarded therefore ‘Single hand upwards’ was awarded six points.

Gestures that had similarities, but the group of academics deemed not similar enough to be awarded a full point, were grouped on a half point basis. For example for the *Reverb + HPF – splash* referent, participant number 7 raises his hand in a ‘wave’ like fashion, and participant number 16 performs a similar action with one of his hands, but on a different axis, while his other hand imitates a knob turn. Therefore, *Hand moves upwards and downwards in a wave motion* was awarded 1.5 points (1 point for number 7’s ‘moves hand in wave like fashion’ and .5 points for the wave movement made by number 16). Figure 14 shows a more complicated example of gesture grouping.

Figure 14. Top: Participant number 4. Bottom: Participant number 5. Although these gestures contain differences they were grouped together in the group: Arm horizontal to arm vertical and back to horizontal (about elbow).

4.7 Agreement scores

The agreement ratings were then calculated for each group. The calculated agreement rates ($AR(r)$) can be found in Table 14. Once the agreement rates were calculated they were then classified using the gesture agreement classification (Vatavu & Wobbrock, 2015) (Table 13). Due to the grouping of similar gestures with .5 scores as opposed to only using full points, numbers that included .5 were plugged into the agreement rate ($AR(r)$) equation.

Table 13. Vatavu and Wobbrock's (2015) gesture agreement classifaction

$AR(r)$	Interpretation
$\leq .100$	*low agreement*
$.100 - .300$	*medium agreement*
$.300 - .500$	*high agreement*
$> .500$	*very high agreement*

4.8 Results

The majority of the referents received low agreement rates. 20 out of 28 resulted in low agreement. 7 out of 28 resulted in medium agreement. 1 out of 28 resulted in high agreement. There we no gestures that resulted in a very high agreement. When using the term consensus throughout the results and analysis sections it refers to gestures that were proposed by two people or more. An overview of the results can be found in Table 14. This table displays the agreement rates ($AR(r)$) and the highest consensus gesture for each referent. All of the consensus gestures are displayed in Table 15 – 23. Images of all of the highest consensus gestures are available in Appendix B; these gestures are shown through five step image overviews. See https://selene.hud.ac.uk/gesturevideo/GestureCompilations/ for videos of all of the elicited gestures.

Table 14. An overview of the GES results, the highest consensus gesture and agreement rates ($AR(r)$) are displayed for each referent

Technique Group	Technique/Referent	Highest Consensus Gesture	$AR(r)$
Filter	HPF - sweep up then down	Arm horizontal to arm vertical and back to horizontal (about elbow)	0.056
Filter	HPF - sweep up	Both hands upwards & Single hand upwards	0.085
Filter	HPF - sweep down	Both hands downwards in a circular fashion	0.092
Filter	LPF - sweep up then down	Hands centre to outstretched to side to centre again	0.033
EQ	EQLow - kill	Single hand downwards	0.029
EQ	EQLow - attenuate	Both hands upwards	0.057
EQ	EQLow - boost	Both hands downwards & Body squat downwards	0.017
EQ	EQLow - vary attenuation	Knob turn metaphor	0.023
EQ	EQMid - kill	Both hands downwards	0.101
EQ	EQMid - attenuate	Both hands downwards	0.114
EQ	EQMid - boost	Both hands upwards	0.040
EQ	EQHigh - kill	Single hand downwards	0.024
EQ	EQHigh - attenuate	Both hands downwards	0.067
EQ	EQ High - boost	Both hands upwards	0.076
Fader	Fader - fade in	Both hands upwards	0.227
Fader	Fader - fade out	Single hand downwards	0.246
Fader	Fader - volume up	Both hands upwards	0.314
Fader	Fader - volume down	Both hands downwards	0.176
Fader	Fader - kill	Single hand downwards	0.111
Fader	Fader – vary	Single hand downwards then upwards	0.086
Effect	Echo + HPF - increase wetness	Little box big box while moving upwards	0.005
Effect	Echo + HPF -increase wetness then decrease delay time	Drumming (holding the sticks) metaphor	0.014
Effect	Reverb + HPF - increase wetness	Little box big box	0.039
Effect	Reverb + HPF - splash	Both hands open and move out as if throwing something then close and move back in	0.019
Loop	Loop - Engage 4 beat	Grabs something picks it up, performs small circle then throws the something	0.012
Loop	Loop - Engage 4 beat then decrease until 1/64th of beat	Shakes closed hand on every loop point & Shakes both hands on every loop point & Big Box little Box, gets smaller on every loop length change	0.014
N/A	Spinback	Vinyl Spinback metaphor	0.179
N/A	Pause Play	Vinyl Spinback metaphor	0.038

Table 15. The filter referent consensus gestures

Filters		
Referent	**Gesture**	**Number of Proposals**
HPF - Sweep up then down	Arm horizontal to arm vertical and back to horizontal (about elbow)	4
HPF - Sweep up then down	Single hand upwards then downwards	3
HPF - Sweep up then down	Both hands upwards then downwards	2.5
HPF - Sweep up then down	Horizontal axis, Right to Left and back	2
HPF - Sweep up	Both hands upwards	4.5
HPF - Sweep up	Single hand upwards	4.5
HPF - Sweep up	Arm horizontal to arm vertical about elbow	3
HPF – Sweep down	Both hands downwards in a circular fashion	5.5
HPF – Sweep down	Knob turn metaphor	3
HPF – Sweep down	Single hand downwards	3
HPF – Sweep down	Arm Vertical to arm horizontal about elbow	2
LPF - Sweep up then down	Hands centre to outstretched to side to centre again	3.5
LPF - Sweep up then down	Single hand downwards then upwards	3

Table 16. The EQLow consensus gestures

EQLow		
Referent	**Gesture**	**Number of Proposals**
EQLow - Kill	Single hand downwards	3
EQLow – Kill	Both hands upwards	2
EQLow – Kill	Both hands downwards	2
EQLow - Kill	Knob turn metaphor	2
EQLow – Attenuate	Both hands upwards	3.5
EQLow – Attenuate	Both hands downwards	2.5
EQLow - Boost	Both hands downwards	2
EQLow - Boost	Body squat downwards	2
EQLow - Vary	Knob turn metaphor	3
EQLow - Vary	Single hand downwards then upwards	2.5

Table 17. The EQMid consensus gestures

EQMid		
Referent	**Gesture**	**Number of Proposals**
EQMid – Kill	Both hands downwards	6
EQMid – Kill	single hand upwards	3
EQMid – Kill	Single hand downwards	2
EQMid – Kill	Both hands upwards	2
EQMid – Kill	Arm vertical to arm horizontal (about elbow)	2
EQMid – Attenuate	Both hands downwards	4.5
EQMid – Attenuate	single hand downwards	3
EQMid – Attenuate	Knob turn metaphor	2
EQMid - Boost	Both hands upwards	3
EQMid - Boost	Full Body Squat Down	2

Table 18. EQHigh consensus gestures

EQHigh		
Referent	**Gesture**	**Number of Proposals**
EQHigh - Kill	single hand downwards	3
EQHigh – Kill	single hand upwards	2
EQHigh - Kill	Both hands upwards	2
EQHigh – Attenuate	Both hands downwards	4
EQHigh – Attenuate	Single hand closes	2
EQHigh - Boost	Both hands upwards	4
EQHigh – Boost	Full Body Squat Down	2
EQHigh – Boost	Knob turn metaphor	2

Table 19. Fader Slow consensus gestures

Fader Slow		
Referent	**Gesture**	**Number of Proposals**
Fader – Fade in	Both hands upwards	8.5
Fader – Fade in	Single hand upwards	6
Fader – Fade in	Knob turn metaphor	2
Fader – Fade out	Single hand downwards	8
Fader – Fade out	Both hands downwards	7
Fader – Fade out	Full Body Squat Down	2.5
Fader – Fade out	Knob turn metaphor	2
Fader – Volume Up	Both hands upwards	10
Fader – Volume Up	Single hand upwards	7
Fader – Volume down	Both hands downwards	7.5
Fader – Volume down	Single hand downwards	5
Fader – Volume down	Full Body Squat Down	3

Table 20. Fader Fast consensus gestures

Fader Fast		
Referent	**Gesture**	**Number of proposals**
Fader - Kill	Single hand downwards	5
Fader – Kill	Freeze	4.5
Fader – Kill	Jazz hands freeze	3.5
Fader – Kill	Grab metaphor	2
Fader - Vary	Single hand downwards then upwards	6
Fader – Vary	Alternate hand chopping	2
Fader – Vary	Fader metaphor	2
Fader – Vary	Single Hand Opens then Closes	2

Table 21. The effect consensus gestures

Effects		
Referent	**Gesture**	**Number of Proposals**
Echo + HPF - increase wetness	Little box big box while moving upwards	2
Echo + HPF - increase wetness then decrease delay time	Drumming (holding the sticks) metaphor	3
Reverb + HPF - increase wetness	Little box big box	4.5
Reverb + HPF - increase wetness	Hands move backwards and forwards constantly reach upwards then back over body	2
Reverb + HPF - splash	Both hands open and move out as if throwing something then close and move back in	2.5
Reverb + HPF – splash	Single hand opens then closes	2
Reverb + HPF - splash	Little box big box	2

Table 22. The loop consensus gestures

Loop		
Referent	**Gesture**	**Number of Proposals**
Loop - engage 4 beat	Grabs something picks it up, performs small circle then throws the something	2
Loop - engage 4 beat then decrease until 1/64th of beat	Shakes closed hand on every loop point	2
Loop - engage 4 beat then decrease until 1/64th of beat	Shakes both hands on every loop point	2
Loop - engage 4 beat then decrease until 1/64th of beat	Big Box little Box, gets smaller on every loop length change	2

Table 23. The Vinyl Spinback and Play/Pause consensus gestures

No Group		
Referent	**Gesture**	**Number of Proposals**
Vinyl Spinback	Vinyl Spinback metaphor	6.5
Play/Pause	Vinyl Spinback metaphor	3
Play/Pause	dancing freeze	2

4.9 Analysis

4.9.1 Agreement ratings

The highest agreement ratings can all be found in the fader referents. One reason for this may be that the fader referents are the least abstract. Gestures such as waving a hand up or down to inform someone to turn the volume up or down are woven into our cultures (Jensenius et al., 2010). Thus it is fair to assume that the experiment was more likely to elicit these familiar gestures resulting in higher agreement ratings.

4.9.2 Conflicting and inconsistent gestures

There were three types of conflicting and inconsistent gestures, these are explained in detail below.

The conflicts are gestures that gained the highest consensus across referents from the same and different groups. The main example of this is the *(both) hand(s) up/down* gestures that gained the highest consensus among a number of the EQ referents and across EQ, volume and filter referents. *(both) hand(s) up/down* gestures refers to the following gestures: *Single hand up*, *Both hands up*, *Single hand down* and *Both hands down*. If a gesture set was designed with these conflicting gestures then the *(both) hand(s) up/down* would control several audio processes which would result in weak action-sound relationships. This could create confusion among the audience as they would see the DJ raising their hand in order to raise the volume and change the cut off of a filter.

The first type of inconsistency were gestures that were proposed between test subjects for a specific referent. For example, *EQLow – kill*, where three participants elicited a *single hand* *down* gesture and two participants elicited a *both hands up* gesture. This could clearly cause confusion among the audience as one would assume that hands moving in the opposite direction would execute the reverse of the process e.g. raising ones hands boosts the *EQLow* then lowering ones hands attenuates the *EQLow*.

The second type of inconsistency were gestures that were proposed between test subjects across referents in the same group. For example, the highest consensus gestures for both *EQLow – Kill* and *EQHigh – Kill* was *single hand downwards*. Berndt et al. (2016) found that these types of inconsistencies were present in their elicited gestures. They stated that in order to create a gesture set they must be removed; suggesting that the overall gesture could stay similar but minor alterations to the movements would have to be made in order to differentiate between the referents. The conflicting and inconsistent gestures are discussed throughout this project.

4.9.3 Detailed Gesture Analysis

4.9.3.1 Referents

EQMid – Kill and *EQMid – Attenuate* both received a medium agreement rating. When compared to the other EQ referents they received a significantly higher agreement rating, but *EQMid – boost*, received a low agreement rating. One explanation for the significantly higher agreement ratings for *EQMid – Kill* and *EQMid – Attenuate* could be that the *EQMid* changes were less subtle than the other EQ referents; the participants may not have had access to headphones/speakers that could reproduce the *EQ Low* or *EQ High* frequencies. Therefore, the test subject had a higher chance of creating a more reliable mental model, which could result is a greater number of identical gestures. This theory is supported by the fader referents, as these referents were most easily understood by all participants, which resulted in high agreement ratings for all fader referents.

The vinyl Spinback referent gained a medium agreement rating. Although the sonic output could be classified as abstract, the vinyl Spinback technique produces a unique sound that has featured on a number of Hip Hop and EDM tracks and is thus widely recognised. The sound producing action of a DJ placing his hand onto a vinyl record/CDJ platter and pulling it backwards, is a widely known gesture; this can be clearly seen when people pretend to DJ they often mimic the action. Therefore, due to the strong action-sound relationship, it is fair to assume that the participants were more likely to produce a metaphorical gesture. This agrees with previous findings within HCI research, as Chan et al. (2016) state that during their GES metaphorical gestures tend to have a higher agreement.

The EQ and filter referents received consensus for the lowering/raising an arm about the elbow style gestures which may suggest that some of the participants associated the filtering with the EQs. This is the case because filtering is a type of EQ.

A number of the inconsistencies are present across the EQ referents. An example of this can be seen for *EQLow – kill*, where three participants elicited a *single hand down* gesture and two participants elicited a *both hands up* gesture. This type of inconsistency can be found among all three EQ bands. It had been anticipated that removing/attenuating frequencies would elicit a downwards movement and boosting frequencies would elicit an upwards movement. This assumption is based on the authors mental model, where removing frequencies lowers the overall output of a signal, and would thus elicit a lowering motion, and boosting frequencies raises the overall output of a signal, thus eliciting a raising motion.

This mental model is most likely influenced by graphical user interfaces on music technology equipment; the user raises the fader to increase the volume and lowers the fader to decrease the volume. This mental model is similar to the hands moving upwards/downwards that was mentioned for the fader referents which can be seen to have strong ties to Western culture. Therefore, this movement may not map to volume but to intensity in general. Balin and Loviscach (2011) results support this argument, as their most popular gesture is a stroke up/down gesture to control intensity.

Several of the consensus gestures for the EQ referents can also be found in the fader referents. This was to be expected as EQ manipulations are frequency dependant volume changes, meaning that when the EQ is changed the overall level of the signal is also altered. This point was mentioned several times throughout the tests, as the majority of participants who had no experience with music technology struggled to understand the EQ referents and some even referred to them as "funny volume changes".

One notable observation for the fader slow referents is that the consensus gestures are all performed along the vertical axis. This was expected as traditional music technology and DJ interfaces tend to control volume through vertical line faders that move upwards and downwards. This is echoed throughout all the proposed gestures apart from one outlier; participant 10 for *Fader – Volume down* proposed a hands inwards along the horizontal axis.

When comparing the consensus gestures from the slow fader referents to the fast fader referents, two factors stand out. Firstly the slow fader referents gained much higher consensus for the *(both) hand(s) up/down* gestures (Average number of proposals for the fast fader referents = 14.75, Average number of proposals for the slow fader referents = 6). The defining factor between these two groups is the speed at which the changes occur. The sudden changes that can be heard within the fast fader referents make lowering and raising ones hand a laborious task. This is also reflected in the smaller range of the raising and lowering of hands within the fast fader referents. It is likely that the significant speed difference between the two groups of gestures forced participants to create completely different mental models, even though the parameters were the same. This is reflected in the number of consensus gestures, the fast fader referents gained four consensus gestures, as opposed to the slow fader referents which gained seven consensus gestures. Secondly, as opposed to the fader slow consensus gestures which were all performed on the vertical axis, for the fast fader changes there were several more horizontal axis based gestures proposed, this could also be a result of the change in mental model that was mentioned above.

The direction inconsistencies that can be seen among the EQ referents are not present in the fader referents. Moving upwards increased the volume and moving downwards decreased the volume. This could be a result of the awareness of gestures that raise/lower the volume that are embedded within Western culture.

The effect based referents received little consensus. The largest group was *little box big box* which received four and a half people for *Reverb + HPF – Increase Wet Dry*. This is clearly a result of the abstract nature of this type of referent. The gestures proposed for the effect based referents tend to be a more flamboyant/animated than those proposed for the less abstract referents, a simple explanation for this can be the gestures reflect the abstract nature of the referents. Another explanation may be that due to the complicated nature of the sonic output, the participants were not aware of what was occurring in the referents, therefore they proposed random movements that did not fit any particular mental model. The latter is least likely due to the universal awareness of the reverb and echo effects.

The ‘Pause Play’ referent received a *Vinyl Spinback metaphor* gesture. This is because the stopping of the track with ‘Vinyl break mode’ active sounds similar to the *Vinyl Spinback* referent.

4.9.3.2 Outliers

During the GES a single participant (participant number 2) proposed 4 gestures for all of the referents which were all metaphorical. These gestures were a clear indication that this participant had been influenced so heavily by his past experience as a DJ that he could not be removed from his normal flow of thinking; even though he was encouraged to do so several times throughout the test. This is an example of legacy biasing. Although a number of other participants proposed such gestures, they did so for two or less referents, therefore participant number 2 was an outlier.

4.9.3.3 Live environment vs Home Environment

It was anticipated that participants would propose different gestures in their home environments as opposed to the club environment. However, there was only one major difference where the participants who performed the elicitation from their homes tended not perform full body gestures, which resulted in a gestures which were performed by their arms and hands exclusively. Six out of the nine participants who performed the GES in the live environment proposed full body gestures however only one of the home environment participants proposed fully body gestures (participant number 21). This finding can be attributed to the lack of room that the majority of participants had in their home environments. This had little impact on the results as the consensus gestures tended to be made up of arm/hand only gestures.

4.9.3.4 Music vs White Noise

The gestures proposed for the white noise referents were similar to the types of gestures proposed for the track based referents. This proves that the specific tracks did not influencing the participants. However, the pulsed white noise emulated the rhythmical elements that are found in the tracks. This resulted in participants proposing gestures that were clearly biased by the pulse. For example, participant 13 for the *Echo + HPF – increase wetness* referent, who clicks their fingers on every pulse. The white noise referents were used in an attempt to remove such biasing therefore the white noise referents proved that the rhythmical elements distract participants during audio processing GES; resulting in biased gestures.

4.9.4 Issues Presented by the Elicited Gestures

A problem that can be found among the EQ referents is the inability to signify which EQ band the respective gesture controls. The participants felt that the same gesture should be used for all EQ bands. However, this is unintuitive and could confuse members of the audience. At the end of the study some of the participants mentioned that they thought this may be a problem and proposed that their gestures should be performed relative to their bodies. Either following a vertical scale or a horizontal scale. The latter proposal replicates the standard music production EQ paradigm, where the frequency rises from low to high; left to right respectively. Therefore, hand up/down gestures would be performed respective to the performers body: left for the EQ Low, centre for the EQ Mid and right for the EQ High. Due to the fact that the participants were not informed of the referents before the task the participant could not take this proposal into consideration during the elicitation study. This highlights a negative of performing the standard GES. To solve this problem the gestures could be assigned to groups prior to the randomisation. In this case the participants would elicit gestures for each type of process in blocks.

The conflicts across the volume, EQ and filter referents are problematic. When implementing a gesture-based system, each referent needs a different gesture. Additionally, if the same gesture was used to control a number of different referents this would result in a weak action-sound relation as when the DJ performed the gesture the audience members would not be aware of which audio processing the gesture would execute.

A number of reasons could explain this phenomenon. The EQ based referent were seen as 'funny volume changes' by several of the non-music-tech experienced participants, this may explain why the high consensus gestures from both of these categories conflict. Another explanation for the conflicting gestures is that by raising and lowering their hands, the participants were simply controlling the intensity of the process, past literature supports this (Balin & Loviscach, 2011). Regardless of the cause of this problem, it is important to remove these conflicts in order to create a gesture that has strong action-sound links.

4.10 Conclusion

The only referents that did not receive low consensus were the fader referents, the vinyl Spinback and two of the EQ Mid referents. Of these gestures only one (*Fader – Volume up*) received high consensus. On one hand the majority of gestures received low agreement but all referents received at least one gesture with some consensus (at least two people proposed the same gesture). Therefore, an end-user DJ gesture set can be defined from the GES featured in this chapter. However, it is important to recognise that the majority of the gestures received low consensus so one could claim that this is not an authentic end-user gesture set as it does not accurately represent a high number of users' mental models.

5 – Gesture Set Creation 1

The results from the GES contained conflicting gestures across the referents. Therefore, a methodology was required that would allow participants to assign a single gesture to a single referent; forcing the participants to assign the conflicting gestures to only one referent would remove the conflicts. Balin and Loviscach (2011) provide a gesture mapping methodology that facilitated such a process. During this study 30 touch gestures, and 22 DAW tasks (referents) were presented via an online survey and test subjects were instructed to assign one gesture to one of the referents. A drawback to using this method is that the participants could assign a gesture to a referent which it was not originally elicited from. However, in an attempt to remove the conflicting gestures an adaptation of this methodology was used for this experiment.

The test was hosted online due to Covid-19 lockdown restrictions. Figure 15 shows an annotated image of the test interface. Eleven of the original audio referents were deleted for this study because they were similar to one of the remaining referents. This resulted in seventeen unique audio referents that were displayed on the left of the interface. Thirty consensus gestures (i.e. gestures that received a score of 2 or more) that were elicited from these seventeen referents were displayed on the right of the interface. The participants were instructed to match one of the videos to one of the referents. To access the application visit https://selene.hud.ac.uk/u1562554/questionnaire/test.php

The test had an attrition rate of 87.5%. This high attrition rate, which is much higher than the average attrition rate found in lab experiments (Walton & Evans, 2018) could be attributed to time required to map a selection of 30 gesture videos to 17 audio referents.

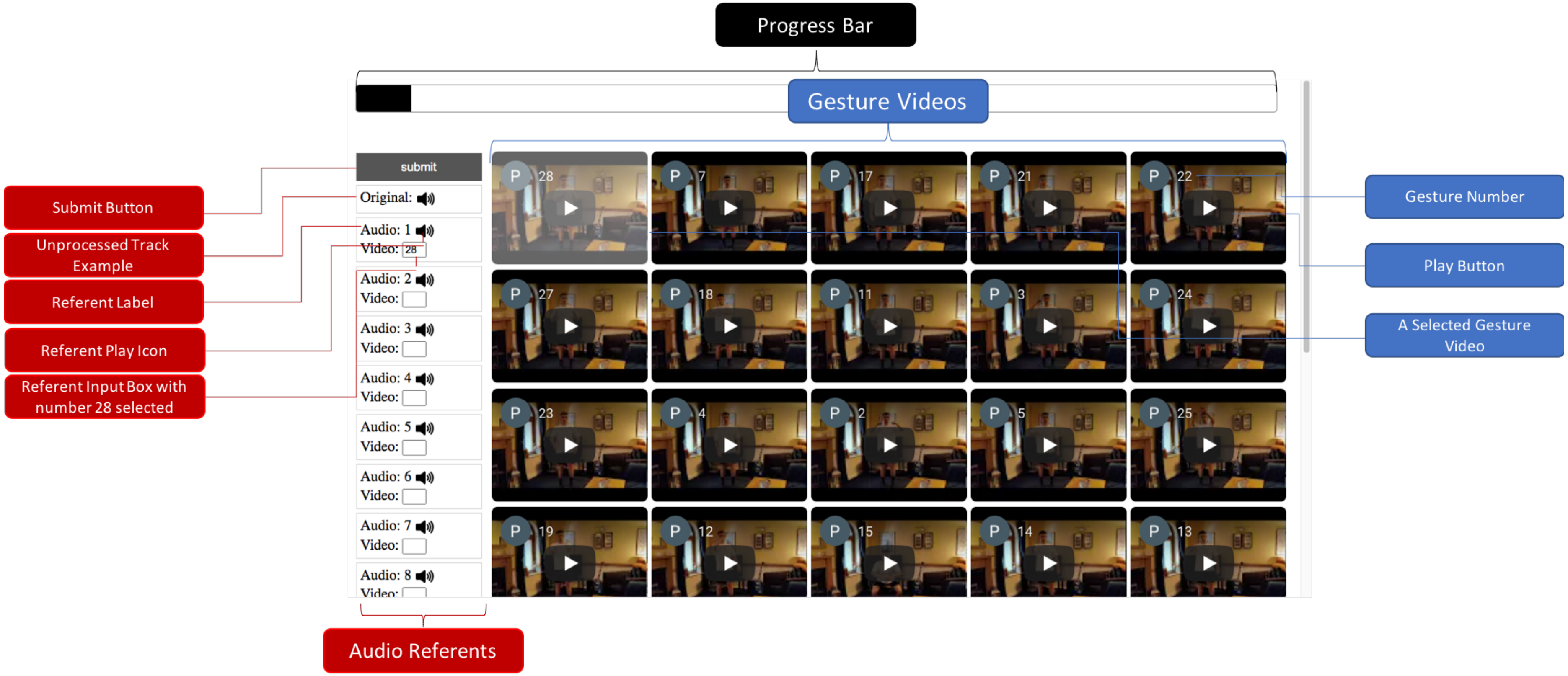


Figure 15. An annotated image of the GSC1 test interface

Unfortunately this study was unsuccessful in removing the original inconsistencies in the GES results. The only audio referents that received a significant number of mappings for the same gesture (i.e. 5 participants or more) were *Fader – Fade out, Loop – engage then half until 1/64th of a beat, LPF – Sweep Down* and *Vinyl Spinback* (see https://selene.hud.ac.uk/gesturevideo/GCSRawData.xlsx for the full results). Therefore, another experiment was devised to try to create a gesture set that does not contain any inconsistencies.

6 - Gesture Set Creation 2

The results from the GSC1 experiment contained inconsistencies which were prominent throughout the EQ referents. During GSC1 an experimental method was adopted which allowed the participants to map any gesture to any referent, regardless of which referent the gesture had been elicited from. This goes against traditional gesture evaluation methodologies and potentially caused confusion. Therefore, the Gesture Set Creation 2 (GSC2) experiment was designed which adopted a more traditional approach in the aim of creating a gesture set which did not contain the aforementioned inconsistencies. Furthermore, by using a more traditional approach this experiment could also act as a form of validation; if the highest consensus gestures from the GES were also most preferred in this study, then they are validated. An adaptation of Morris et al. (2010) methodology is adopted for this experiment.

6.1 Experiment Design

Again this experiment had to be performed online due to the ongoing Covid-19 lockdown. The approach outlined by Morris et al. (2010) was adapted to facilitate online testing (see section 2.6.4 for details of this methodology). The subjective rating stage of the methodology was removed to reduce the total test time. The participants simply had to select which gesture they thought was most suitable for each referent.

For each referent the participant was shown three or four gesture videos in 2 x 2 grid on the same screen. The videos were time synchronised and played back simultaneously. The Wizard-of-Oz technique was adopted so that the gestures in each video appeared to be controlling the audio processing. To decrease the time required to complete the test the number of audio referents was reduced. Table 24 displays the audio referents that were removed from the study with a brief description as to why this decision was made. Morris et al. included Wobbrock et al.'s (2009) gestures for comparison purposes; these gestures acted as a benchmark (anchors). Due to the minimal research that has been performed for eliciting gestures for audio processing including anchors in this study was not feasible. Berndt et al.'s (2016) gestures were not used as anchors as the study biased the participants by using sensing hardware and introducing them to a pre mapped gesturally controlled system prior to the elicitation.

Table 24. The selection criteria of the audio referents for GSC2

Referent Name	**Was it utilised during this study?**	**Reason for removal**
HPF - Sweep up then down	Yes	N/A
HPF - Sweep up	No	These two gestures are sub components of *HPF – Sweep up then down*
HPF - Sweep down	No	
LPF - Sweep up then down	Yes	N/A
EQLow – Kill	Yes	N/A
EQLow - Attenuate	No	These two gestures are sub components of *EQLow - Kill*
EQLow – Boost	No	
EQLow - Vary attenuation	Yes	N/A
EQMid – Kill	Yes	N/A
EQMid - Attenuate	No	These two gestures are sub components of *EQMid - Kill*
EQMid – Boost	No	
EQHigh – Kill	Yes	N/A
EQHigh - Attenuate	No	These two gestures are sub components of *EQHigh - Kill*
EQ High - Boost	No	
Fader - Fade in	Yes	N/A
Fader - Fade out	Yes	N/A
Fader - Volume up	No	'Fader – fade in' performed this over a longer period of time and was more popular during the identification of DJ techniques.
Fader - Volume down	No	'Fader – fade out' performed this over a longer period of time and was more popular during the identification of DJ techniques.
Fader – Kill	Yes	N/A
Fader – Vary	No	This is the same as multiple repetitions of *Fader - Kill*
Echo + HPF - increase wetness	Yes	N/A
Echo + HPF - increase wetness then decrease delay time	Yes	N/A
Reverb + HPF - increase wetness	Yes	N/A
Reverb + HPF - splash	Yes	N/A
Loop - Engage 4 beat	No	This is a simple version of *Loop - Engage 4 beat then decrease until 1/64th of beat*
Loop - Engage 4 beat then decrease until 1/64th of beat	Yes	N/A
Spinback	Yes	N/A
Play/Pause	Yes	N/A

6.1.1 Gesture set criteria

Table 25 displays all of the selected gestures.

Table 25. All of the gestures for each referent displayed in the order they appear on the grid for GSC2

Referent	Gesture A	Gesture B	Gesture C	Gesture D
HPF – Sweep up then down	Single hand upwards then downwards	Horizontal axis, right to left then back	Both hands upwards then downwards	Arm horizontal to arm vertical and back
LPF – Sweep up then down	Hands centre to outstretched and back	Arm vertical to arm horizontal and back	Full body downwards then upwards	Single hand downwards then upwards
EQLow - Kill	Both hands downwards	Both hands upwards	Single hand downwards	Knob turn metaphor
EQLow - Vary	Knob turn metaphor	Single hand fist pumps above head	Single hand opens then closes	Single hand downwards then upwards
EQMid – Kill	Arm vertical to arm horizontal	Both hands downwards	Both hands flip	Single hand upwards
EQHigh – Kill	Single hand flips	Single hand upwards	Single hand downwards	Both hand upwards
Fader – Fade in	Both hands upwards	Single hand upwards	Knob turn metaphor	Single hand opens
Fader - Fade out	Single hand downwards	Full body squats downwards	Knob turn metaphor	Both hands downwards
Fader – Kill	Jazz hands freeze	Single hand downwards	Freeze [holds hand out]	Freeze [large 'X' with both arms]
Echo + HPF - increase wetness	Single hand swipes to the side	Little box big box	Little box big box while moving upwards	One hand up and down on every echo + the other hand upwards over duration
Echo + HPF - increase wetness then decrease delay time	Hand outstretched in front moves inwards	Little box big box whilst moving upwards	Hands outstretched move inwards whilst moving backwards and forwards constantly	Both hands drumming
Reverb + HPF - increase wetness	Hands reach behind body whilst moving backwards and forwards constantly	Both hands raise to either side into Jesus pose	Little box big box on every pulse	Little box big box
Reverb + HPF - splash	Little box big box	Single hand opens then closes	Both hands open and move out as if throwing something	Hand in front of chest moves outwards
Loop - Engage 4 beat then decrease until 1/64th of beat	Hand closed to open	Shakes both hands on every loop point	Shakes single hand on every loop point	Big box little box
Spinback	Knuckles together move outwards	Full body spins round in a circle	Vinyl Spinback metaphor	N/A
Play/Pause	Hand moves downwards as if pressing down on something in the air	Dancing freezing	Vinyl Spinback metaphor	N/A

The gestures that received the highest consensus during the GES were chosen for this study. If a referent received more than four consensus gestures, the two highest consensus gestures were selected. The two remaining gestures were selected based on suitability. A number of referents did not receive four consensus gestures therefore the same selection process was used as mentioned above. The author then video recorded himself performing all of the selected gestures. The consensus gestures were made up of groups of similar gestures that contained minor differences. Therefore, when recording, the best representation of all of the gestures in the group was created.

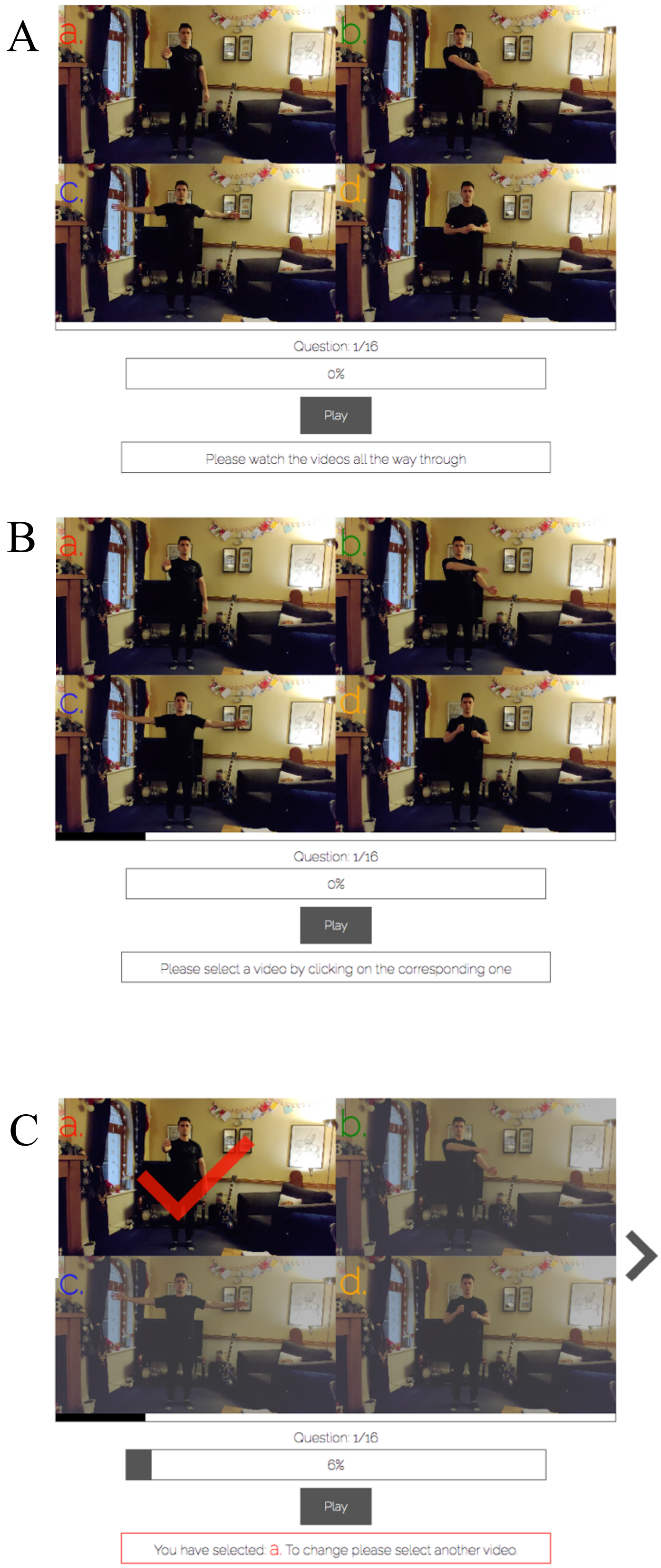


Figure 16. The GSC2 interface in three states. A) the participant is instructed to watch to watch the video before they can move on B) once the video has been viewed the participant is instructed to select their preferred gesture before they can move on C) feedback is given as to which video has been selected, the participant can now move onto the next referent

6.1.1.1 Interface

Figure 16 shows three images of the interface in the different states. To select their preferred gesture the participant simply clicks on the respective video. The application was designed and developed to be accessed on laptop/desktop PC browsers.

6.2 Testing

Given the successes in recruiting participants via Reddit, this social media website was used solely to promote the test. The test was promoted on the following threads: acidtechno, DJs, DnB, EDM, electronicmusic, gabber, hardstyle, hci, House, liquiddnb, minimal, NativeInstruments, tech_house, Techno, trance and trap.

The participants were given a chance to win a £50.00 Amazon voucher. Prior to performing the test, participants were presented with a questionnaire (see https://selene.hud.ac.uk/u1562554/idstudy2/). The test was uploaded to the University of Huddersfield, School of Computing and Engineering's server, to view the test please visit: https://selene.hud.ac.uk/u1562554/idstudy2/test

6.3 Results

125 participants were recruited. Due to technical issues three of the participants' details were not logged. 114 were male and 8 were female. The ages of the participants ranged from 10 – 60; the largest age group was 16-30 year olds whom 85 of the participants belonged to. 59 of the participants were experienced DJs. 106 of the participants were experienced audience members of dance music events. 37.5% of the participants progressed from the questionnaire

to the activity. Of this 37.5%, 28% finished the test. Therefore the attrition rate for this test was 72%. This is 15.5% lower than the attrition rate for the previous test which is an improvement but it is still much higher than the average in lab test attrition rate (Walton & Evans, 2018).

Throughout the analysis the referents are split into two categories. Simple audio processing referents and complex audio processing referents. The referents that constitute the categories are detailed in the Table 26.

Table 26. The simple audio processing and complex audio processing referent groups

Simple Audio Processing Referents	**Complex Audio Processing Referents**
1. HPF - Sweep up then down	10. Echo + HPF - Increase wetness
2. LPF - Sweep up then down	11. Echo + HPF - Increase wetness then decrease delay time
3. EQLow – Kill	12. Reverb + HPF - Increase wetness
4. EQLow - Vary attenuation	13. Reverb + HPF – Splash
5. EQMid – Kill	14. Loop - Engage 4 beat then decrease until 1/64th of beat
6. EQHigh – Kill	15. Vinyl Spinback
7. Fader - Fade in	16. Play/Pause
8. Fader - Fade out	
9. Fader – Kill	

Two statistical methods were utilised to analyse the results. Firstly a Chi squared goodness of fit test was performed on each referent to determine whether there was a preference, these results are shown in Table 27. The null hypothesis (H_1) is equal to “there was no preference”, and the alternative hypothesis (H_a) is equal to “there was a preference”. The chosen level of significance (α) was 0.05.

Table 27. The GSC2 Chi squared goodness of fit results

Referent	**Chi Squared Value (χ^2)**	**Degrees of freedom (K)**	**Critical Value**	**Hypothesis**
1. HPF - Sweep up then down	7.96	3	7.815	H_a
2. LPF - Sweep up then down	29.656	3	7.815	H_a
3. EQLow – Kill	23.128	3	7.815	H_a
4. EQLow - Vary attenuation	31.768	3	7.815	H_a
5. EQMid – Kill	18.648	3	7.815	H_a
6. EQHigh – Kill	79.896	3	7.815	H_a
7. Fader - Fade in	26.712	3	7.815	H_a
8. Fader - Fade out	19.096	3	7.815	H_a
9. Fader – Kill	13.08	3	7.815	H_a
10. Echo + HPF - Increase wetness	29.528	3	7.815	H_a
11. Echo + HPF - Increase wetness then decrease delay time	81.56	3	7.815	H_a
12. Reverb + HPF - Increase wetness	104.794	3	7.815	H_a
13. Reverb + HPF – Splash	117.656	3	7.815	H_a
14. Loop - Engage 4 beat then decrease until 1/64th of beat	80.792	3	7.815	H_a
15. Vinyl Spinback	55.744	2	5.991	H_a
16. Play/Pause	43.888	2	5.991	H_a

Then averages were calculated along with 95% confidence intervals. Figure 17 - 20 display histograms of the percentage of votes that each gesture received for all of the referents.

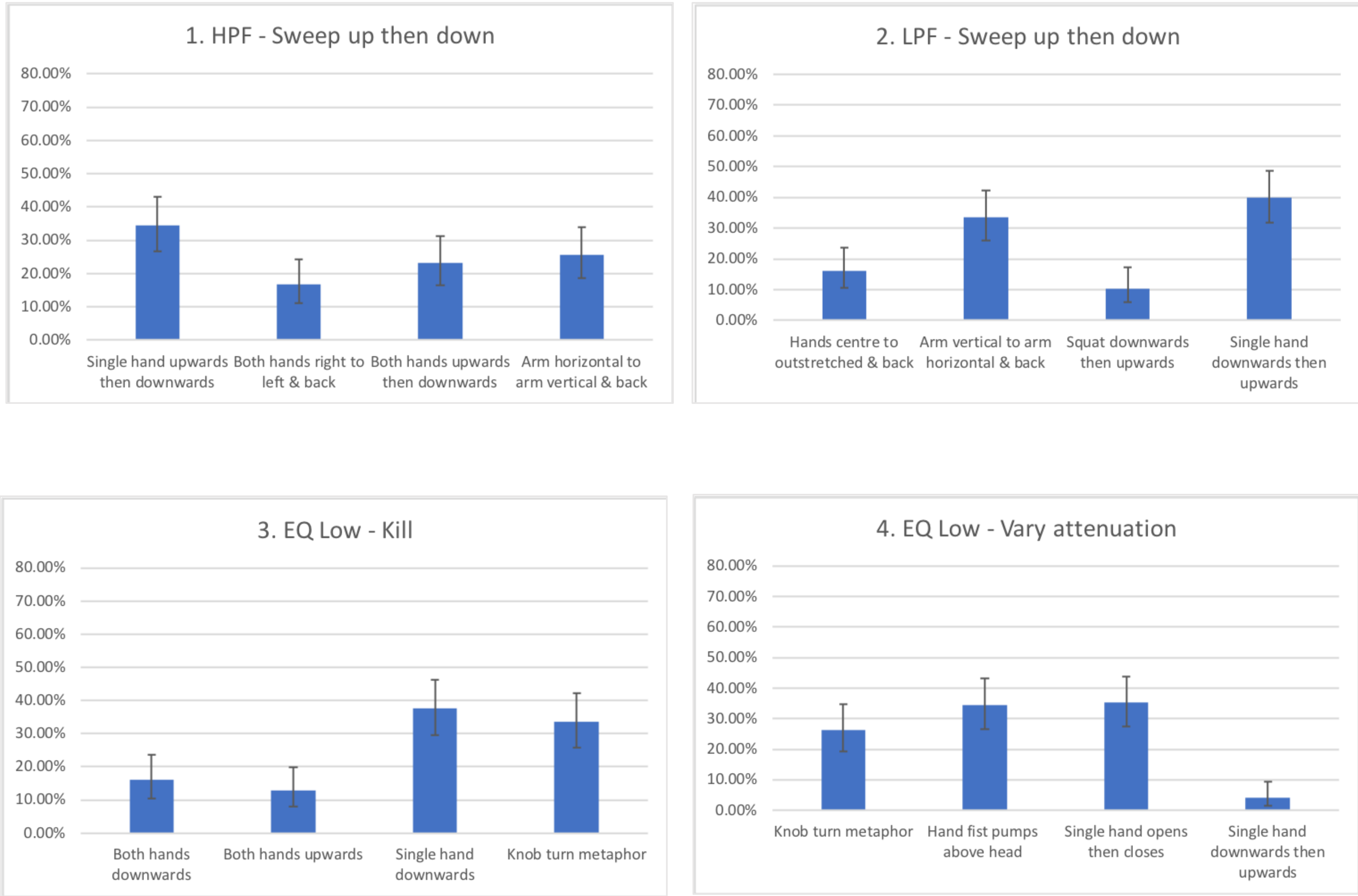


Figure 17. Average scores with 95% confidence intervals for 1) HPF - Sweep up then down 2) LPF - Sweep up then down 3) EQLow - Kill 4) EQLow – Vary

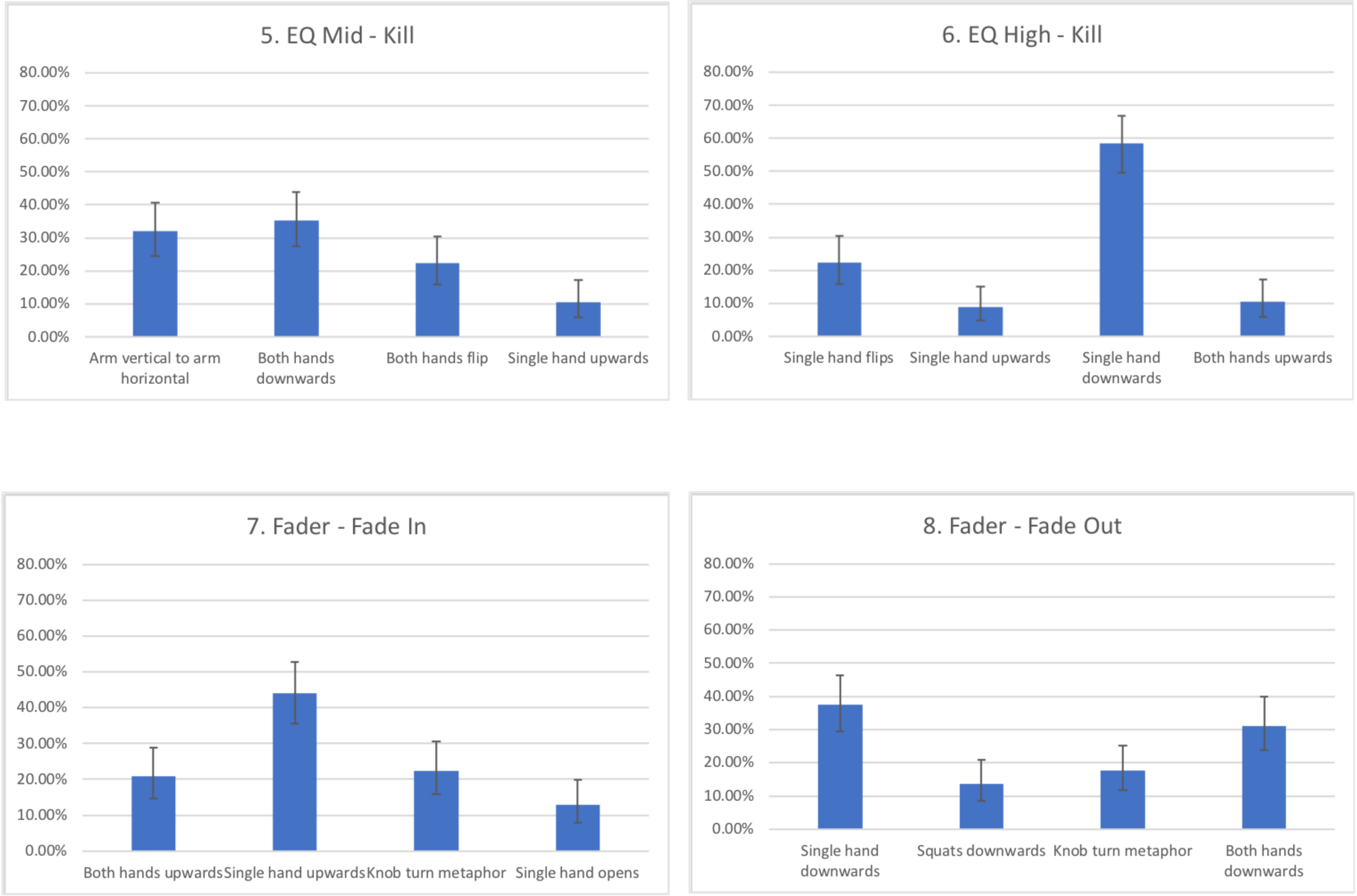


Figure 18. Average scores with 95% confidence intervals for 5) EQMid – Kill 6) EQHigh – Kill 7) Fader – Fade in 8) Fader – Fade out

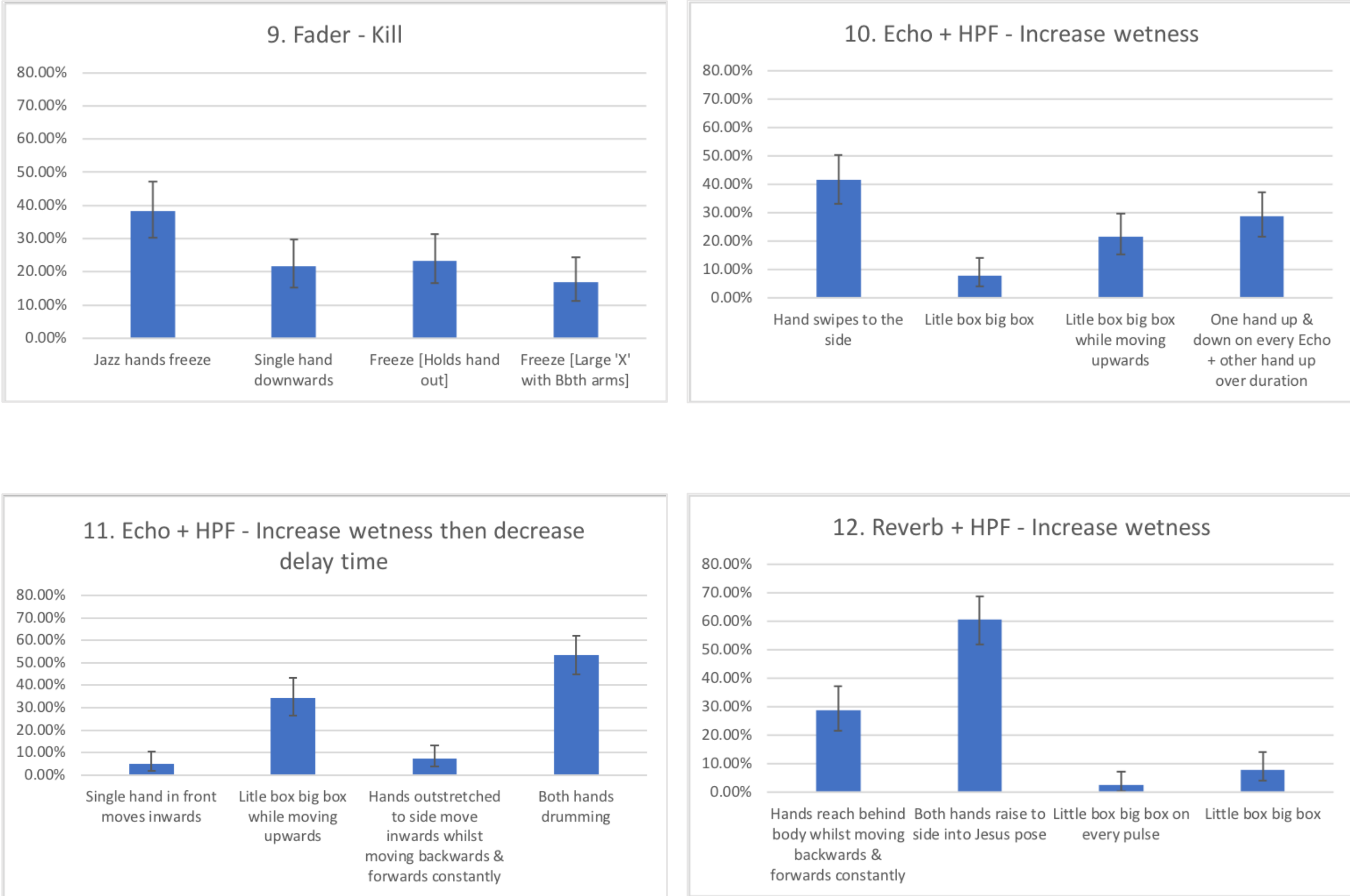
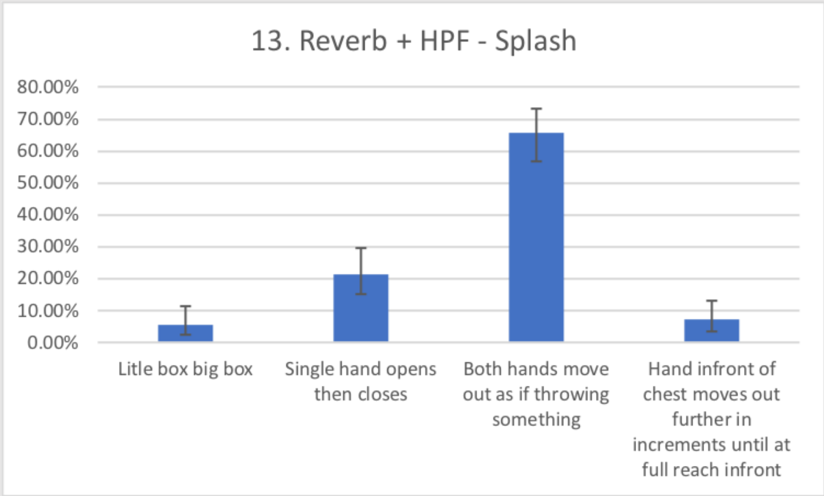
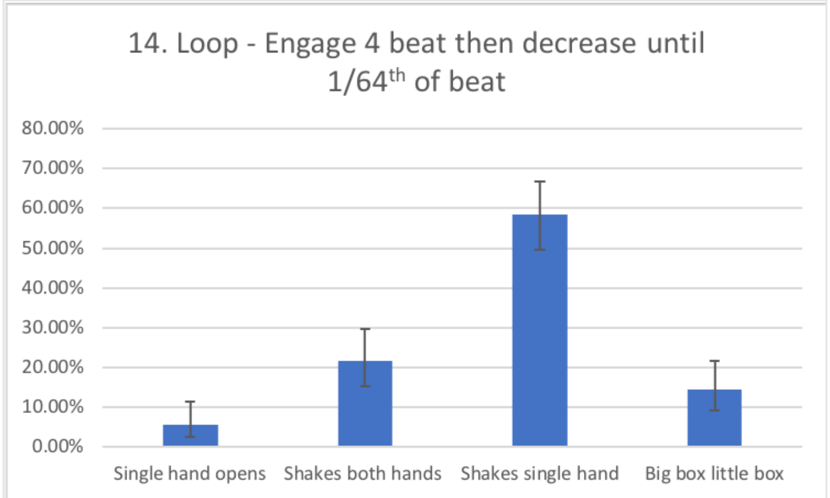
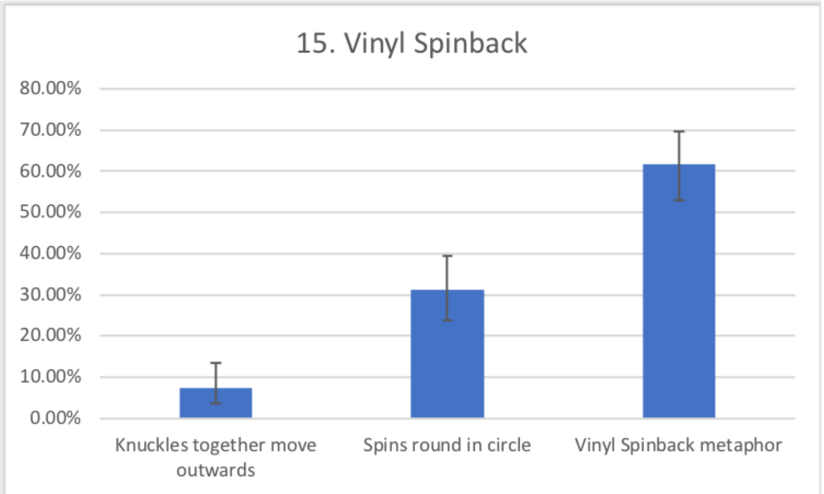


Figure 19. Average scores with 95% confidence intervals for 9) Fader – Kill 10) Echo + HPF – Increase wetness 11) Echo + HPF – Increase wetness then decrease delay time 12) Reverb + HPF – Increase wetness







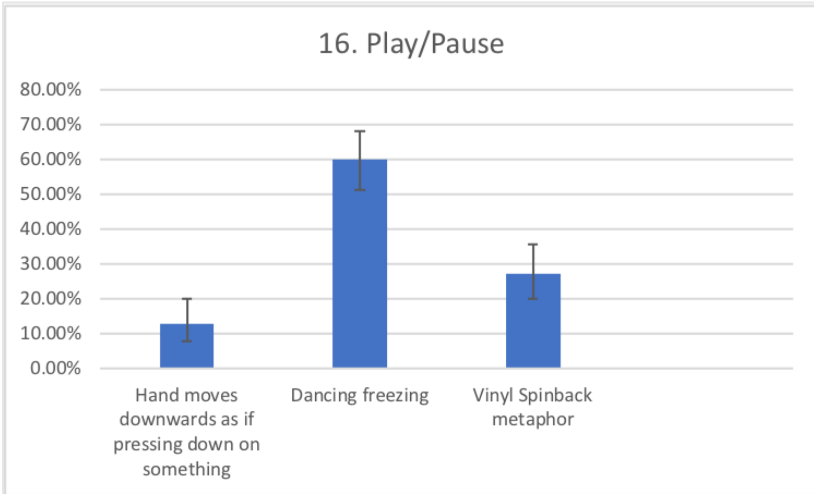


Figure 20. Average scores with 95% confidence intervals for13) Reverb + HPF – Splash 14) Loop – Engage 4 beat then decrease until 1/64th of a beat 15) Vinyl Spinback 16) Play/Pause

6.4 Analysis of results

Table 28 shows the highest percentage score from this study along with the highest consensus gesture from the GES for comparison.

Table 28. Shows the gesture with the highest percentage score from GSC2 and the highest consensus gesture from the GES for each audio referent.

Group	Audio Referent	Gesture with highest percentage score	Is there a distinct preference (i.e. non-overlapping confidence intervals)?	Is it the highest consensus gesture in the GES?	Score for this gesture in the GES
Simple audio processing	1. HPF - sweep up then down	Single hand upwards then downwards	No (only against fourth gesture)	No	3
Simple audio processing	2. LPF - sweep up then down	Single hand downwards then upwards	No (only against third & fourth gesture)	No	3
Simple audio processing	3. EQLow - kill	Single hand downwards	No (only against third & fourth gesture)	Yes	3
Simple audio processing	4. EQLow - vary attenuation	Single hand opens then closes	No (only against fourth gesture)	No	1
Simple audio processing	5. EQMid - kill	Both hands downwards	No (only against fourth gesture)	Yes	6
Simple audio processing	6. EQHigh - kill	Single hand downwards	Yes	Yes	3
Simple audio processing	7. Fader - fade in	Single hand upwards	Yes	No	6
Simple audio processing	8. Fader - fade out	Single hand downwards	No (only against third & fourth gesture)	Yes	8
Simple audio processing	9. Fader - kill	Jazz hands freeze	No (only against third & fourth gesture)	No	3.5
Complex audio processing	10. Echo + HPF - increase wetness	Hand swipes to the side	No (only against third & fourth gesture)	No	1
Complex audio processing	11. Echo + HPF -increase wetness then decrease delay time	Both hands drumming	Yes	Yes	3
Complex audio processing	12. Reverb + HPF - increase wetness	Both hands raise to the side into Jesus pose	Yes	No	1
Complex audio processing	13. Reverb + HPF - splash	Both hands move out as if throwing something	Yes	Yes	2.5
Complex audio processing	14. Loop - engage 4 beat then decrease until 1/64th of beat	Shakes single hand	Yes	Yes	2
Complex audio processing	15. Vinyl Spinback	Vinyl Spinback metaphor	Yes	Yes	6.5
Complex audio processing	16. Play/Pause	Dancing freezing	Yes	No	2

Throughout this analysis two statistical tests have been performed. Firstly, a Chi squared goodness of fit test was performed to define whether there was a preference for any gesture. All of the Chi squared values were greater than the critical value, which means that the alternative hypothesis (H_a) was met. The alternative hypothesis (H_a) is equal to "there was a preference", therefore all of the referents received a preference. However, this test does not define whether a specific gesture was preferred.

Inspection of the histograms in Figure 17 - 20 suggest that the complex audio processing referents received a distinct preference (i.e. non-overlapping confidence intervals) whereas the simple audio processing referents did not. The direction inconsistencies that were prominent during GSC1 were not present in these results, as the EQ Kills highest scores were downwards motion gestures and the filter sweeps received downwards and upwards motions which aligned with the sweeping of the filter. However, the *(both) hand(s) up/down* conflicting gestures that were present across the EQ, filter and volume referents during the GES are present in these results. Five gestures received a distinct preference for gestures that had received the highest consensus during the GES, therefore these gestures were validated and could be classed as true end-user gestures.

From the complex audio processing group *Echo + HPF – Increase WetDry And then ½ delay time*, *Reverb + HPF – Splash*, *Loop – Engage four Bar then ½ until 1/64 of a beat*, and *Vinyl Spinback* all received a distinct preference for a gesture during this experiment which was the same as their highest consensus gesture from the GES. These results validated the high consensus gestures. These gestures are all metaphorical which tend to be understood by audience members easier (Brown et al., 2018) and receive higher consensus during elicitation studies (Chan et al., 2016). Furthermore, the metaphorical gestures that received high consensus during the GES (which were featured in the gestures sets of this study) have strong action-sound relationships. For example, the *Vinyl Spinback Metaphor* gesture mimics a DJ rubbing their hand on a vinyl record; this gesture is easily translated to this action which is widely known to produce the sonic output. Therefore, these referents likely received a distinct preference due to the metaphorical nature of their gestures.

Only one of the simple audio processing referent (*EQHigh – Kill*) received a distinct preference that validated the high consensus gesture from the GES. The data suggests that the *(both) hand(s) up/down* gestures are preferred for *HPF – Sweep up then down*, *LPF – Sweep up then down*, *EQLow – Kill*, *EQMid – Kill*, *EQHigh – kill*, *Fader – Fade In* and *Fader – Fade Out*. This finding supports the earlier claim that stated the *(both) hand(s) up/down* gestures are not intended to be mapped to a single parameter but to the intensity of audio processors in general. Past literature also supports this argument as Balin and Loviscach (2011) found that their highest score was a stroke up/down gesture that was mapped to intensity. However, the mapping of a single gesture to all of the EQ, volume and filter processors appears unintuitive. If in one instance a DJ raised their hand(s) and an HPF swept upwards, then seconds later they raised their hand(s) and the volume increased, this could lead to weak action-sound relationships. Furthermore, such gestures could further degrade the existing DJ-audience communication problem. This form of interaction is analogous to the DJ turning a knob, which could be controlling any number of parameters, on traditional hardware. Therefore, a method of removing these conflicts is required.

Finally, it is also important to note that in many cases the test subjects preferred gestures in GSC2 that do not match the highest consensus gesture in the GES. For example, the *HPF – Sweep up then down* referent gained the highest score for *Single hand upwards then downwards* during this experiment however this referent's highest consensus gesture from the GES was *Arm horizontal to arm vertical and back to horizontal (about elbow)*. These findings support Choi et al.'s (2012) claim that end-users do not necessarily prefer high consensus gestures from GES. Furthermore, when performing GES in the future subjective ratings or expert opinions should be used in conjunction with the GES.

6.5 Conclusion

Every referent received Chi squared goodness of fit scores that state there was a preference. Five referents (*EQMid- Kill*, *Echo + HPF – Increase WetDry And then ½ delay time*, *Reverb + HPF – Splas*h, *Loop – Engage four Bar then ½ until 1/64 of a beat*, and *Vinyl Spinback*) were validated as they received a distinct preference. This was unexpected as a number of these referents, especially the complex audio processing referents, received lower consensus during the GES. Therefore, one could claim that the complex audio processing referents have been validated and a resulting gesture set could be created from the high preference gestures. In contrast, the simple audio processing referents tended not to receive a distinct preference. The inconsistencies that were prominent in the EQ referents from the GES and GSC1 were not present during this experiment. However, the simple audio processing referents contained the *(both) hand(s) down/up* conflicting gestures which were discovered during the GES. The conflicting gestures cannot be used in a DJ gesture set due to the weak action-sound relationship that such gestures would create. Furthermore, a final study is performed in order to create a conflict free gesture set.

7 - Gesture Set Creation 3

The aim of this experiment was to create an end-user gesture set for DJ performance. However, due to the lack of consensus among the elicited gestures and the volume, filter and EQ referent conflicts, a set could not be defined from the GES or both of the Gesture Set Creation experiments. This is the third attempt at creating a gesture set after the issues with the first two attempts. A final experiment was designed that evaluated gesture sets, which were created from the results from GSC2 with the aim of outputting a conflict free end-user gesture set. This would result in a gesture set that has unique gestures for each audio process.

7.1 Experimental Design

7.1.1 Methodology

An adaptation of Morris et al. (2010) evaluation study methodology is used again during this experiment. The premise of this methodology is showing participants videos of gestures who then rate these gestures based on suitability. However, Morris et al. were testing the suitability of individual gestures, whereas this experiment aimed to test whole gesture sets. Therefore, videos of whole gesture sets would be displayed as opposed to videos of single gestures. This experiment aimed to compare and validate gesture sets which were developed from the GSC2 preference gestures.

The audio referents from this study were divided into the same groups that were used during the analysis of the previous section: simple audio processing referents and complex audio processing referents. This was performed due to the clear divide in the results of GSC2, in which the simple audio processing referents tended not to receive a gesture with a distinct preference whereas the complex audio processing referents tended to receive a gesture with a distinct preference. Therefore, four gesture sets were designed for the simple audio processing referents, but only two gestures sets were designed for the complex audio processing referents. Each group contained an anchor gesture set. The gesture sets developed are shown in Table 29 and Table 30; the following sections go into details on how they were designed.

Table 29. The manufactured gesture sets for the simple audio processing referents

Audio Referent	Gesture Set 1	Gesture Set 2	Gesture Set 3	Gesture Set 4 (anchor)
1. HPF – Sweep up then down	Arm horizontal to arm vertical about elbow	Arm horizontal to arm vertical about elbow	Arm horizontal to arm vertical about elbow	Hand Outstretched in front of body palm facing forwards, move arm inwards to body
2. LPF – Sweep up then down	Single hand downwards then upwards	Arm vertical to arm horizontal about elbow	Arm vertical to arm horizontal about elbow	Hand Open to Claw
3. EQ Low – Kill	Knob turn metaphor	Knob turn metaphor	Hand Down [Left]	Scrubbing hands together
4. EQ Low – Vary	Single hand opens then closes	Single hand opens then closes	Single hand opens then closes	Sideways swipe
5. EQ Mid – Kill	Arm vertical to arm horizontal about elbow	Both hands downwards	Hand Down [Middle]	Spin around 180°
6. EQ High – Kill	Single hand downwards	Single hand downwards	Single hand downwards [Right]	Lean Backwards
7. Fader – Fade In	Single hand upwards	Single hand upwards	Both hands upwards	Fader Metaphor {small}
8. Fader – Fade Out	Both hands downwards	Squats Down	Both hands downwards	Fader metaphor {small}
9. Fader – Kill	Jazz hands freeze	Jazz hands freeze	Jazz hands freeze	Raise hand

Table 30. The manufactured gesture sets for the complex audio processing referents

Audio Referent	**Gesture set 5**	**Gesture set 6 (anchor)**
1. Echo + HPF - Increase WetDry	Swipe to the side	Single hand held out in front of chest static
2. Echo + HPF - Increase WetDry and ½ delay time	Both hands drumming	Hand moves around in circle
3. Reverb + HPF – Increase WetDry	Raise hands to the side in Jesus pose	Spin around in circle
4. Reverb + HPF – Splash	Hands move out as if throwing something	Throws either arms to the side about shoulders
5. Loop – Engage 4 bar loop then ½ until 1/64th of a beat	Hand shakes	Single hand closes
6. Vinyl Spinback	Vinyl Spinback	Knob turn metaphor
7. Pause Play	Dancing freeze	Flips both hands about elbow

7.1.2 Simple Audio Processing Gesture Set Design

7.1.2.1 Simple Audio Processing Gesture Set 1

The basis of the first set was created by selecting all of the most preferred gestures from GSC2. This resulted in a set that contained a number of conflicting gestures (Table 31).

Table 31. The basis of gesture set 1

Referent	**Gesture**	**Conflicting?**
1. HPF – Sweep up then down	Single hand upwards then downwards	Yes
2. LPF – Sweep up then down	Single hand downwards then upwards	Yes
3. EQ Low – Kill	Single hand downwards	Yes
4. EQ Low – Vary	Single hand opens then closes	No
5. EQ Mid – Kill	Both hands downwards	Yes
6. EQ High – Kill	Single hand downwards	Yes
7. Fader – Fade In	Single hand upwards	Yes
8. Fader – Fade Out	Single hand downwards	Yes
9. Fader – Kill	Jazz hands freeze	No

(Both) Hand(s) Up/Down gestures gained the highest score for various audio referents. To remove these conflicts, the audio referent that received the highest score for the *(Both) Hand(s) Up/Down* gesture in GSC2 was assigned this gesture, the remaining audio referents

that contained the conflicts were assigned their second highest scored gesture from GSC2. This was performed in descending order, starting with the highest scores and then descending to the lowest scores. For example, *EQ High – Kill* gained a score of 73 (the highest result for all non-abstract based referents so this was performed first) for *Single hand down*. Therefore, *EQ Low – Kill* and *Fader – Fade out* were assigned their second highest consensus gesture, i.e. *Knob turn metaphor* and *Both hands downwards* respectively. This process was followed until all of the audio referents received a unique gesture. The resulting gesture set can be found in Table 32.

Table 32. Intermediate version of gesture set 1

Referent	**Gesture**	**Conflicting?**
1. HPF – Sweep up then down	Single hand upwards then downwards	Yes with *Fader – Fade in*
2. LPF – Sweep up then down	Single hand downwards then upwards	Yes with *Fader – Fade out*
3. EQ Low – Kill	Knob turn metaphor	No
4. EQ Low – Vary	Single hand Opens then closes	No
5. EQ Mid – Kill	Arm vertical to horizontal about elbow	No
6. EQ High – Kill	Single hand downwards	No (this received the highest score)
7. Fader – Fade In	Single hand upwards	Yes with *HPF – Sweep up then down*
8. Fader – Fade Out	Both hands downwards	Yes with *LPF – Sweep up then down*
9. Fader – Kill	Jazz hands freeze	No

This resulted in a set that still contained conflicting gestures. *HPF – Sweep up then down*, *LPF – Sweep up then down*, *Fader – Fade in* and *Fader – Fade out* all contained the conflicting *(both) hand(s) up/down* gestures. The gestures for the *Fader – Fade in* and *Fader – Fade out* are subcomponents of the gestures for the *HPF – Sweep up then down* and *LPF – Sweep up then down* so this clearly will not work as a gesture set. The aim of this experiment is to create a gesture set which does not contain the conflicts; each audio process is required to have its own gesture. Therefore, due to the fact that *Fader – Fade in* received a higher score in GSC2 (55 vs 43) *HPF – Sweep up then down* fell back onto its second highest scored gesture. A similar conflict occurred between *LFP – Sweep up then down* and *EQ High – Kill*. *LFP – Sweep up then down* had been assigned *Single hand downwards then upwards* and *EQ High – Kill* had been assigned *Single hand downwards*. For this instance *EQ High – Kill* gained a higher score (73 vs 50). If one was to follow the same method that was performed to solve the conflict between *HPF – Sweep up then down* and *Fader – Fade in* then *EQ High – Kill* would have been assigned the gesture. However, *LFP – Sweep up then down* secondary gesture had already been assigned to *EQMid - Kill* so therefore *EQ High – Kill* was awarded its secondary gesture and *LFP – Sweep up then down* was assigned *Single hand downwards then upwards*. The resulting gesture set can be seen in the second column of Table 29.

7.1.2.2 Simple Audio Processing Gesture Set 2

The second set was designed to favour the referents that had received a stronger preference for the highest scored gestures during GSC2. As opposed to removing the conflicts by awarding the highest scoring gesture in GSC2 to the audio referent, the gestures which had the largest difference in score between the highest scored gesture and the second highest scored gesture in GSC2 was used (Table 33). This was performed in descending order, starting at the audio referent which received the largest difference, then descending until at the audio referent with the lowest difference.

Table 33. The simple audio processing highest scoring gesture and second highest scoring gesture with the difference in scores

Referent	Highest Scoring Gesture	Score	Second Highest Scoring Gesture	Score	Difference
1. HPF	Single hand upwards then downwards	43	Arm horizontal to arm vertical about elbow	32	11
2. LPF	Single hand downwards then upwards	50	Arm vertical to arm horizontal about elbow	42	8
3. EQ Low - Kill	Single hand downwards	47	Knob turn metaphor	42	5
4. EQ Low – Vary	Single hand opens then closes	44	Hand fist pumps above head	43	1
5. EQ Mid – Kill	Both hands downwards	44	Arm vertical to arm horizontal about elbow	40	4
6. EQ High – Kill	Single hand downwards	73	Single hand flips	28	45
7. Fader – Fade In	Single hand upwards	55	Knob turn metaphor	28	27
8. Fader – Fade Out	Single hand downwards	47	Both hands downwards	39	8
9. Fader - Kill	Jazz hands freeze	48	Holds hand out	29	19

EQHigh – Kill clashed with *Fader – Fade Out* and *EQLow - Kill*. *Fader – Fade Out* was assigned its fourth highest scored gesture, as the second and third highest scored gestures were already taken by other referents. *Fader – Fade Out* third highest scored gesture was *Knob turn metaphor* which was free, however *Knob turn metaphor* was *EQLow - Kill* second highest scored gesture which received a greater score than *Knob turn metaphor* did for *Fade – Fade Out*. Therefore, *Knob turn metaphor* was assigned to *EQLow – Kill*. The resultant gesture set is shown in the table below (Table 34).

Table 34: Intermediate version of gesture set 2

Referent	Gesture	Conflicting?
1. HPF – Sweep up then down	Single hand upwards then downwards	Yes with *Fader – Fade in*
2. LPF – Sweep up then down	Single hand downwards then upwards	Yes with *Fader – Fade out*
3. EQ Low – Kill	Knob turn metaphor	No
4. EQ Low – Vary	Single hand opens then closes	No
5. EQ Mid – Kill	Both hands downwards	No
6. EQ High – Kill	Single hand downwards	Yes with *LPF – Sweep up then down*
7. Fader – Fade In	Single hand upwards	Yes with *HPF – Sweep up then down*
8. Fader – Fade Out	Squats down	No
9. Fader – Kill	Jazz hands freeze	No

Again this resulted in similar conflicts. *HPF – Sweep up then down* conflicted with *Fader - Fade in*. *Fader - Fade in* gained the largest difference therefore *HPF – Sweep up then down* was assigned its second highest scored gesture. *LPF – Sweep up then down* conflicted with *EQHigh – Kill*. *EQHigh – Kill* received the largest difference therefore *LPF – Sweep up then down* was assigned its second highest scored gesture. The second gesture set can be found in the third column of Table 29.

7.1.2.3 Simple Audio Processing Gesture Set 3

The third set was designed by the author and both supervisors in the aim of creating an intuitive gesture set. A requirement was set by the author that related processing should have similar gestures that are based on the gestures from GSC2. For example, both of the filter referents should utilise a similar gesture. This was to aid in creating the optimum action-sound coupling gestures. Firstly, the gestures with the highest score were chosen (Table 31). A method of removing the *(both) hand(s) up/down* conflicting gestures was then required. The filter referents contained such conflicting gestures. However, the filters second highest scored gestures, the *Arm bend about elbow* style gestures, received a score of ten less than the highest scored gestures and were both similar style gestures. Therefore, the filter referents were both assigned their second highest scored gestures.

The remaining conflicts were between the EQ and volume referents. Single hand down/up was assigned to all of these referents apart from *EQ Mid – Kill* which was assigned *Both hands downwards*. Due to the similarity requirement, it seemed logical to assign *Single hand downwards* to the EQ kills. To differentiate between the EQ bands the standard music production EQ paradigm was utilised in which frequencies run from left (low) to right (high). Therefore, the *Single hand down* gesture would be performed relative to the performers body: left for EQ low, centre for EQ mid and right for EQ high. This mapping appeared intuitive as lots of DJs are aware of this paradigm which would potentially align with their own mental models. To align with the music production EQ paradigm the fader referents were assigned both hands up/down, as this mimicked moving all of the EQ bands up/down which would have the same effect as raising or lowering the volume. Not only did this make logical sense but both of the fader referents received a significant score in GSC2 for *both hands up/down* gestures so there was some popularity for these gestures. All of the remaining audio referents that didn't have any conflicts remained the same (*EQLow - Vary and Fader - Kill*). The third gesture set can be found in the fourth column of Table 29.

7.1.2.4 Simple Audio Processing Gesture Set 4

An anchor set was included in this experiment in order to test whether the manufactured gesture sets were better than simply randomly allocating gestures. If gesture 1, 2 or 3 received lower subjective ratings than gesture set 4 (the anchor set) then one could state that the respective gesture set is no better than a randomly selected gesture set. The anchor set consisted of non-consensus gestures which were picked at random from the GES. The anchor gesture set can be found in the fifth column of Table 29.

7.1.3 Complex Audio Referent Gesture Set Design

Due to the fact that the majority of the complex audio processor referents received distinct preferences, only two gesture sets were created; a set of all of the highest scored gestures from GSC2 and an anchor set which was created using the same method as was used to create the simple audio processing anchor set (set number 4). Both of these gesture sets can be found in Table 30.

7.1.4 Interface Design

Figure 21 shows an annotated screenshot of the interface for this experiment.

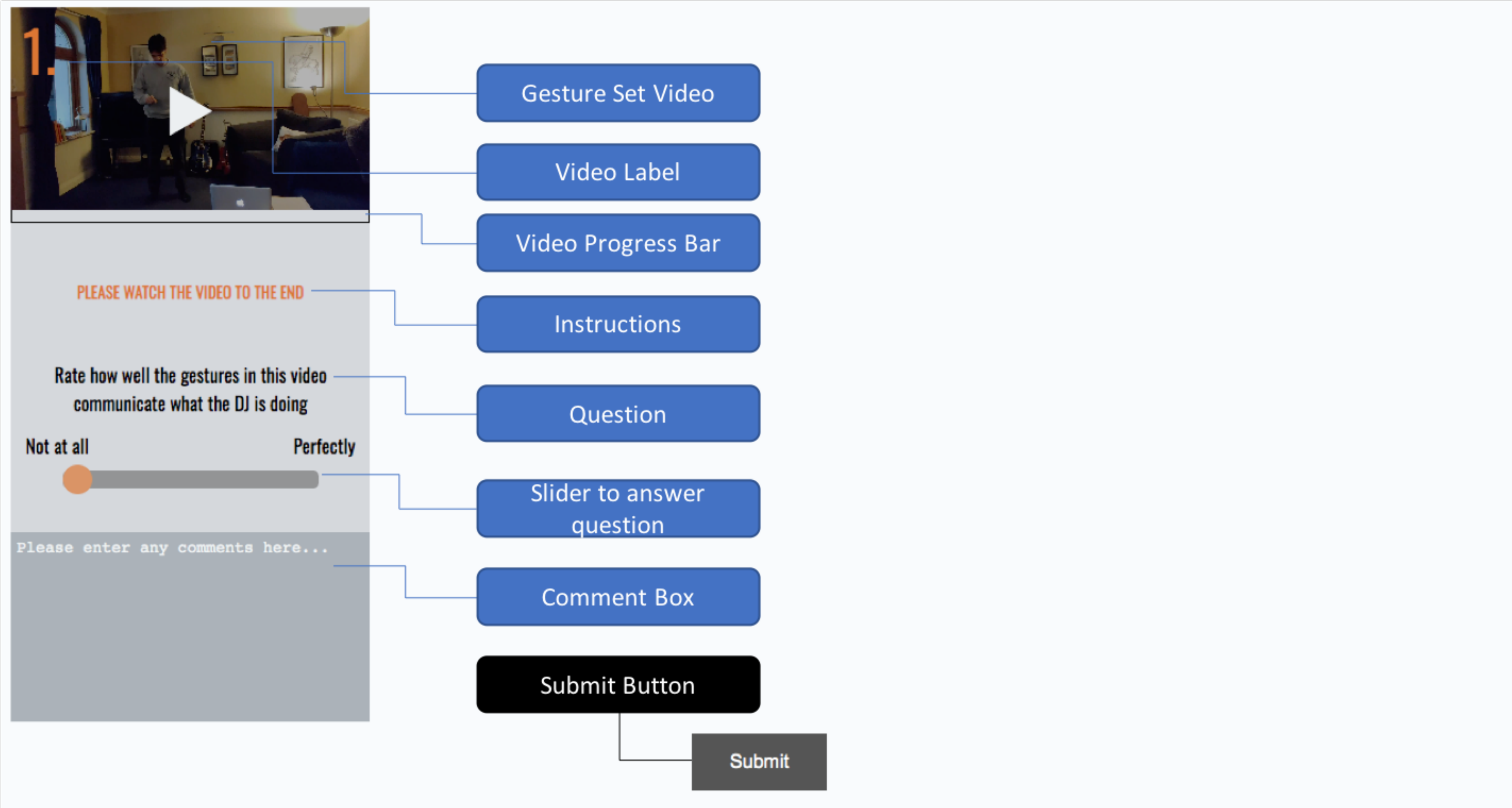


Figure 21. An annotated screenshot of the interface for GSC3

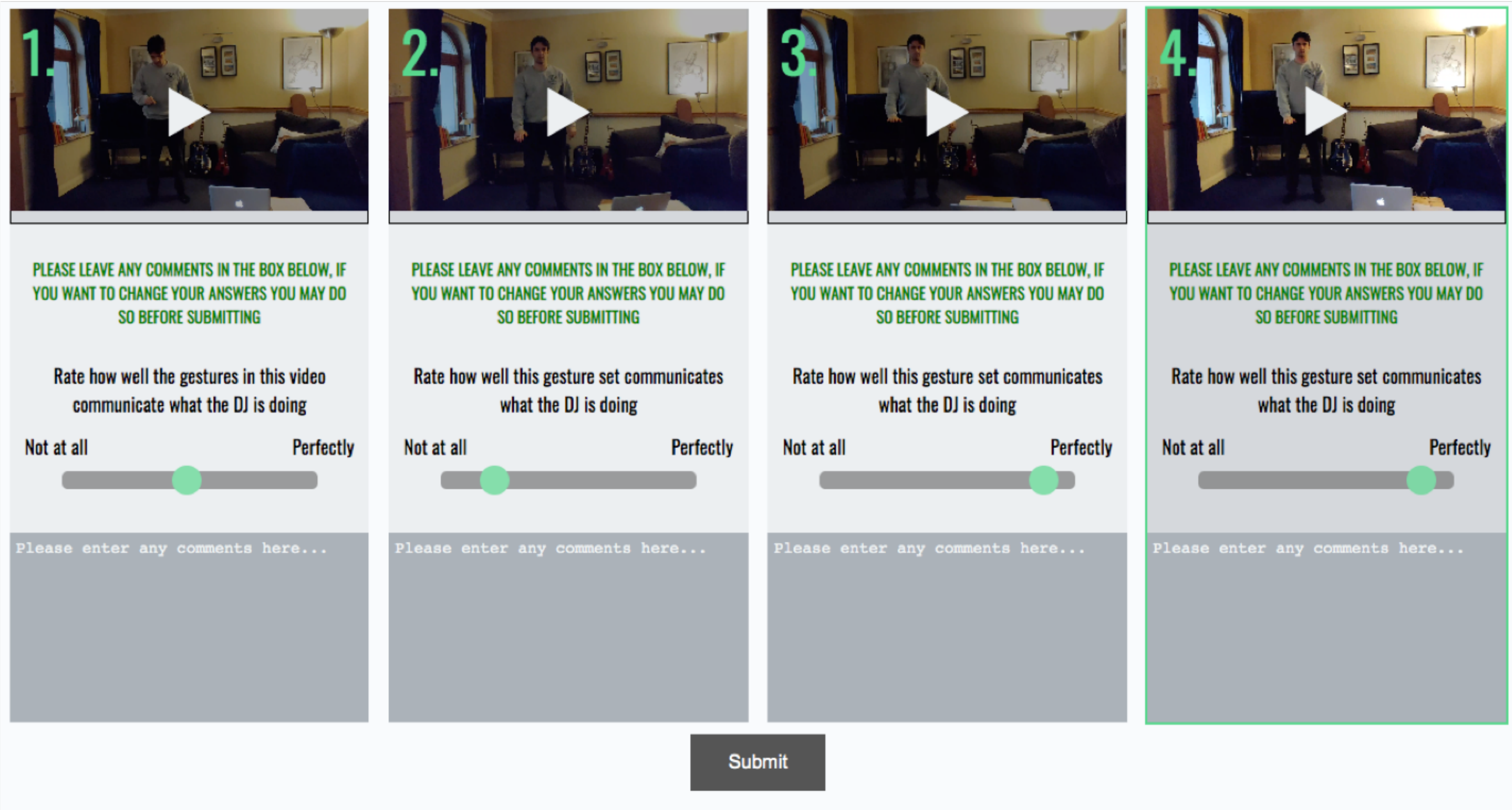


Figure 22. A screenshot of the interface for GSC3 once the participant has viewed all of the simple audio processing gesture set videos

The gesture set videos were displayed in their own independent section. Only one section was displayed upon entering the test, then once the participant had viewed the gesture set video for the current section, the next section appeared (Figure 22).

7.2 Testing

The participants were given a chance to win a £50.00 Amazon voucher. Prior to performing the test, participants were presented with a questionnaire (https://selene.hud.ac.uk/u1562554/GestureSetStudy/). Participants were recruited mainly through reddit but music-tech students and lecturers also took part. The test was uploaded to the University of Huddersfield, School of Computing and Engineering's server, to view the test please visit: https://selene.hud.ac.uk/u1562554/GestureSetStudy/test

7.3 Results

31 participants took part in the test. Due to technical issues two of the participants' details were not recorded. 25 of the participants were male and 4 were female. The age range of the participants was 16-45. The largest number of participants (21) fell into the 16-30 category. 16 of the participants were DJs. 22 participants were recruited from Reddit and 7 participants were music technology students/lecturers. Two participants mentioned that they were having a problem with synchronisation; there is no way of telling whether this issues was more widespread or just isolated to these two individuals. The results were plotted as histograms. Initial visual observation revealed the data was not normally distributed for each gesture set. Therefore, the box plot (Figure 23) was produced to analyse the results.

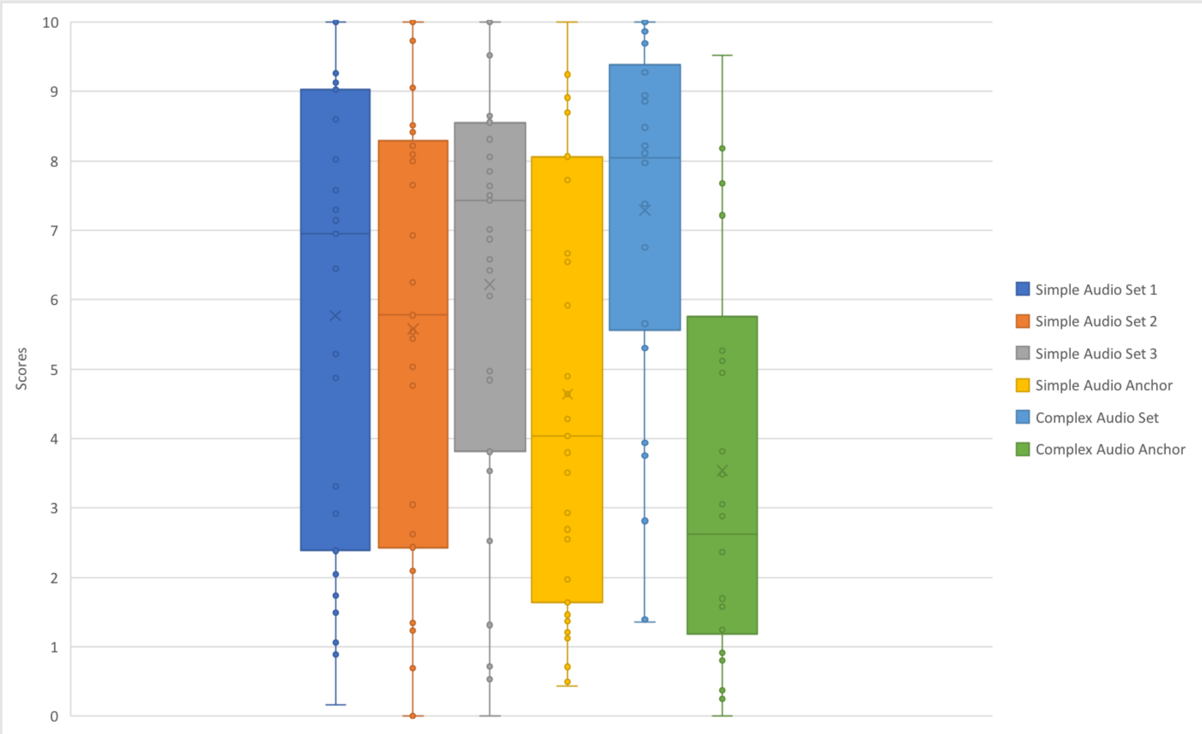


Figure 23. A box plot of the results for GSC3

7.3.1 Analysis of Results

During the box plot analysis set 4 is referred to as the simple audio anchor. Set 1, 2, and 3 have a similar median, however the median for set 3 is greater than the median for both set 1 and 2, suggesting that set 3 is preferred over set 1 and 2. There is a large difference between the median of set 3's and the simple audio anchor's median. The significance of this difference is shown by the fact that the set 3's lower quartile is almost level with the simple audio anchors' median. None of the four simple audio processing boxes state any significant results as all of the boxes overlap greatly. The interquartile ranges are all also similar, but set 3's interquartile range is clearly smaller than the interquartile range of set 1, 2 and the simple audio anchor.

During the box plot analysis of the complex audio processing referents set 5 is referred to as the complex audio set and set 6 is referred to as the complex audio anchor. There is a large difference between the complex audio set median and the complex audio anchor median. The complex audio set's lower quartile slightly overlaps with the complex audio anchors upper quartile. Furthermore, the complex audio set is skewed to the top of the plot, whereas the complex audio anchor is skewed to the bottom of the plot. These three observations strongly suggest that the complex audio set is preferred to the complex audio anchor. The interquartile range of the complex audio set is also significantly smaller than the interquartile range of the complex audio anchor which suggest the ratings are less dispersed and thus a greater number of people selected a similar rating.

Following the above visual analysis of the box plots statistical tests were conducted. The assumption of normality was violated for all gesture sets as assessed by a Shapiro-Wilk test of normality presented in Table 35. In light of this, two nonparametric Friedman tests were conducted to assess the differences between the preference ratings for the gesture sets. The null hypothesis in both cases was that there is no difference between the gesture sets. For the simple audio processing tasks the null hypothesis was rejected meaning there was no statistically significant difference between the preference ratings for gesture sets 1 to 4, $\chi2(3) = 5.503$, $p = 0.138$.

Table 35: GSC3 Shapiro-Wilk test of normality

Gesture Set	Statistic	df	Sig.
GS1	0.898	31	0.006
GS2	0.919	31	0.023
GS3	0.893	31	0.005
GS4	0.909	31	0.012
GS5	0.861	26	0.002
GS6	0.887	26	0.008

In contrast, for the complex audio processing referents, the Friedman test which assessed the difference between Gesture Sets 5 and 6 revealed a statistically significance difference between the two sets, $\chi2(1) = 15.385$, $p < 0.05$. Despite there not been a statistically significant difference for the simple audio processing gesture sets, a combined gesture set was produced using set 3 and the statistically significant set 5. This gesture set is shown in Table 36.

Table 36: The final gesture set

Referent	**Gesture**
1. HPF	Arm horizontal to arm vertical about elbow
2. LPF	Arm vertical to arm horizontal about elbow
3. EQ Low - Kill	Single hand downwards [Left]
4. EQ Low – Vary	Singe hand opens then closes
5. EQ Mid – Kill	Single hand downwards [Middle]
6. EQ High – Kill	Single hand downwards [Right]
7. Fader – Fade In	Both hands upwards
8. Fader – Fade Out	Both hands downwards
9. Fader - Kill	Jazz hands freeze
10. Echo + HPF - Increase WetDry	Swipe to the side
11. Echo + HPF - Increase WetDry and ½ delay time	Both hands drumming
12. Reverb + HPF – Increase WetDry	Raise hands to the side in Jesus pose
13. Reverb + HPF – Splash	Hands move out as if throwing something
14. Loop – Engage 4 bar loop then ½ until 1/64th of a beat	Single hand shakes
15. Vinyl Spinback	Vinyl Spinback
16. Pause Play	Dancing freeze

The aforementioned gesture set is the output from all four experiments. Some of the original audio referents were not covered as they were removed to minimise the test duration of the Gesture Set Creation experiments. However, the concepts from the gestures that are defined in the set can be applied to the absent referents.

The conflicts that were found across the volume, EQ and filter referents were removed. This set utilised the *hand down/up* for the EQ referents; these gestures were performed relative to the performers body to differentiate between the EQ bands, this used the music production EQ paradigm. The *Arm moves about elbow* gestures were utilised for the filter referents. *Both hands down/up* were utilised for the volume referents, these gestures aligned with the music production EQ paradigm that was used for the EQ gestures. The complex audio referent gesture set which was defined by selecting the highest preference gestures from GSC2 gained a distinct preference. This set contained referents for *Echo + HPF – Increase WetDry And* *then ½ delay time*, *Reverb + HPD – Splash*, *Loop – Engage four Bar then ½ until 1/64 of a beat*, and *Vinyl Spinback* which were validated during GSC2. These referents and high preference gestures were featured in gesture set 5, therefore these gestures were again validated during this study. This shows that these gestures were preferred over not only random gestures, but the other gestures that were elicited during the GES.

7.4 Conclusion

This study presented gestures sets, that were created using a number of different methods, to both DJs and non-DJs (audience members) who subjectively rated them. The aim of the study was to create a conflict free gesture set. The analysis of a box plot shows that the complex audio processing set was strongly preferred over the complex processing anchor set. The majority of the gestures found in the complex audio processing set had been validated during the previous study therefore this study acted as further form of validation. Furthermore, one could claim that the complex audio processing gestures are optimal end-user gestures. The analysis of the simple audio processing referents suggest that the set that was manufactured by the author and both supervisors was preferred over the anchor set, however these results were not significant. As a result set 3 and the complex audio processing set were used to create a final gesture set that was preferred over randomly selected gestures.

8 – Discussion

There were a number of complications that were stated during the analysis of the elicited gestures. The two main issues being conflicting gestures across the volume, EQ and filter referents, which saw *(both) hand(s) up/down* style of gesture gaining the highest consensus for multiple referents, and the inconsistent directional gestures among the EQ referents, which resulted in participants eliciting gestures that moved in the opposite direction for the same referent and across referents in the same group. As a result, if a gesture set was created from the highest consensus gestures, the set would be unintuitive and could confuse the audience; a gesture set that contains conflicts and inconsistencies could also further degrade the DJ-audience communication. In addition, the majority of the referents received low agreement rates, which suggests that the this GES was unsuccessful.

In contrast one could claim that due to the abstract nature of the referents that it is unfair to compare the agreement rates from this project to the agreement rates from standard HCI GES. The referents used in traditional HCI GES tend to be much simpler and universally understood. If the participants understand the referents more and are accustomed to using them in in everyday life, then creating a mental model is not only going to be easier but there will also be a higher chance that the mental model aligns with other peoples'; resulting in higher agreement. Therefore, when performing GES with abstract referents the agreement rate classification could be adapted to allow for greater freedom. In this sense, one could claim that the low agreement that is found across the majority of the referents during this project is not low. For example, *Echo + HPF – Increase wetness* gained a single consensus gesture from two participants, and 19 other gesture proposals. This audio process is abstract therefore creating a mental model that executes the process would be difficult. Furthermore, by receiving two of the same gesture proposals this could be seen as a positive result and should be investigated in the future.

There were a number of factors that could have caused the low agreement, conflicts and inconsistencies. Firstly non-DJs performed the GES. From a traditional GES perspective this could appear counterintuitive. The GES is a methodology that was grounded in the user centred design approach (Norman, 2013). This method focuses on observing end-users in their natural environment. The important factor is that users who are going to use the product/system that is being developed are put at the centre of the design process. Audience members are not physically going to be using the gesture set, therefore they are not the end-user. This project focused on creating a gesture set that better communicated common DJ techniques to the audience members. Therefore, the design decision to involve the audience members in the GES was made in an attempt to create universally understood gestures, the audience members were not going to be physically using the gestures to execute the audio processes but they were being designed in order to aid the audiences understanding of what the tasks the DJ is performing. Therefore, eliciting the gestures from the audience perspective appeared intuitive.

However, the low consensus among the elicited gestures could be attributed to involving audience members in the GES as these types of participants did not understand several of the referents. This was made clear throughout the GES as a number of participants did not understand some of the EQ referents and became confused when proposing the gestures for these referents. It is unlikely that the participants would be able to create an accurate mental model when in a state of confusion which would result in arbitrary gestures being proposed. Such gestures are not likely to match other people’s gesture proposals which would result in lower consensus. In future DJ focused GES the audience should not be involved. However, the audience should be involved during the evaluation of the gestures. Audience members could be shown videos of the elicited gestures and asked questions such as "how well does this gesture communicate what is happening in the audio" with an aim at identifying the gestures with the best action-sound relationship.

Another factor that could have caused the aforementioned problems is the use of sonic referents. Traditional GES tend to use primitive computer tasks. During the elicitation the participants are usually displayed videos of the primitive computer task being executed and then asked to propose the gesture. No other information is provided. This is a strict process which is performed in order to force the participant to define their own mental model that is unbiased. Such an approach was followed during the GES performed in this project. A referent was presented to the participants and no other information was given unless the participant displayed signs of struggle; in this case the test facilitator would intervene and began a small discussion on what was occurring in the corresponding referent. Therefore, the participants were unaware of what audio processes were going to be present during the experiment so there were not able to compensate for the fact that there were similar referents with subtle differences. For example, there were three different EQ types. A number of participants commented on this problem towards the end of the experiment stating that they wish they would have been told what audio processes were going to be present so that they could have compensated for the various referents that were similar to each other; the participants stated they wanted to propose similar gestures for the similar referents. By removing the non-DJs (audience members) from the GES this problem could be avoided as the participants would all have the required knowledge to understand what the referents are simply by reading either a one word label or a short text description. In this instance the participant could be informed of all of the referents prior to performing the elicitation and could accommodate for the different EQ bands. In an attempt to avoid any further biasing the referents could be placed into groups (e.g. EQ and filters, faders, effects etc…), the referents would be randomised within the groups and then the order of the groups would be randomised. Then prior to the elicitation the test facilitator could give brief details on what the groups contained. The participants would also be more likely to avoid creating direction inconsistencies as they would be more aware of the different types of referents and thus could factor in methods to differentiate between the referents.

Legacy biasing is a problem that could negatively affect all GES. To mitigate the effects of legacy biasing priming was used. Morris et al. (2014) state that such techniques reduce the effects of legacy bias but was this the case for this study? There are no means of direct comparison as all the participants were primed using the same methods however there are a number of legacy biased inspired gestures present in the elicited gestures. Legacy biasing is when a user's new interactions are influenced by their previous experiences. Due to the ubiquitous WIMP system, new interfaces are all threatened with the user's desire to utilise a mouse and pointer to interact with an icon style display. However, for this project this biasing can be seen by movements that traditional DJs make. It is likely that some of these movements, like twisting knobs and pressing buttons, are well known enough that non-DJs are aware of them and can thus provide legacy biased gestures even though they have no experience with pre-existing DJ technology. A small number of examples of these style of gestures can be found throughout the elicited gestures. Participant 2 only proposed gestures that were clearly wholly based off his standard interactions with DJ equipment. Participant 16, who had no DJ experience, proposed a knob turn metaphor for *EQLow - Kill*.

It is important to note that the highest consensus gesture for *Vinyl Spinback* referent was a metaphorical gesture which imitates the sound producing action for the referent – someone placing their hand on a platter and pulling it backwards. This gesture was validated during GSC1, GSC2 and GSC3; clearly displaying the popularity of this gesture. This suggests that a DJ performance gesture set could benefit from such a gesture, due to its strong action-sound coupling it seems logical to adopt such a gesture. On the other hand, gestures that feature users mimicking turning knobs or moving faders lack such an action-sound relationship are ambiguous and could be detrimental to a DJ gesture set. It was inevitable that these types of gestures were going to be elicited however the number of proposals were that minimal as not to cause any concern. The results of GSC2 support this argument as knob turn gestures received some preference but did not receive the most votes for any of the referents. Therefore, the benefit of legacy bias which resulted in the participants proposing metaphorical gestures clearly out ways the negative aspects.

The final problem that could have negatively impacted the GES performed in this project is the use of consensus as the sole measure to define the most suitable gesture. It is thought that the highest consensus gestures are the most suitable due to the fact that the mental models of the highest portion of the participants were similar. However, a number of the preferred gestures from GSC2 were not the highest consensus gestures from the GES, this suggests that the highest consensus gestures are not always the most preferred; these preferences were not statistically significant therefore this finding requires further investigation but past literature supports this argument (Choi et al., 2012; Chen et al., 2018; Wu et al., 2019). These studies state that a round of subjective ratings should be performed in order to define end-user's most preferred elicited gestures. In contrast, five out of the sixteen referents from GSC2 (*EQHigh*

– Kill, Echo + HPF – Increase WetDry And then ½ delay time, *Reverb + HPD – Splash*, *Loop – Engage four Bar then ½ until 1/64 of a beat*, and *Vinyl Spinback*
) validated the highest consensus gesture from the GES with a distinct preference. This suggests that consensus can be a useful measure but should not be used on its own. Therefore, a combination of consensus and subjective ratings should be used to define the most suitable gestures in future GES.

Due to the problems that have been highlighted throughout this discussion the author suggests avoiding using the strict GES methodology when designing gestures for DJs. Instead, a GES can be performed which acts as a gesture creation exercise. During the gesture creation phase the strict GES methodology should be followed but with a number adaptations. Test subjects should be told minor details about the referents prior to the elicitation. The referents will be elicited in groups. Prior to the elicitation the test facilitator should inform the test subjects that they should try remove any past prejudice from their gestures by focusing on listening carefully to the referents and then producing gestures that are innate. Then two different methods could be used to select the most suitable gestures. The first method involves a round of subjective ratings, then the subjective ratings and consensus could be used to select the best gestures. The second method involves a number of experts who select the gestures that would create the most intuitive gesture set; this is similar to the method that was followed during GSC3 in which the author and both supervisors created set 3.

It is very unlikely that experts alone would have been able to design the majority of the gestures in the final gesture set without consultation with end-users. This is because when designing new interactions past experiences often prejudice the designer into creating interactions that are personal and do not necessarily reflect the end-users' mental models. When designing interfaces experts often assume how users will interact with the system however these assumptions are often biased by their past experiences, this mainly occurs subconsciously thus the designer is unaware of the bias they are placing onto the interface.

This research has produced a gesture set, in the following paragraph the author will discuss the suitability of the constituent gestures.

As a whole the author feels that the gesture set makes sense and is a credible solution. The author feels that the *Arm horizontal to arm vertical about elbow* and the *Arm vertical to arm horizontal about elbow* gestures are a good idea, this is likely due the that fact that they imitate a gate opening and closing. It is unlikely that experts would have been able to design such gestures. This highlights the benefit of performing a GES. The use of the hands up/down gestures for the EQs and fader up/down are clearly suitable, as they are universally understood gestures for increasing the intensity of a number of parameters. The metaphorical gestures that appear in the set are clearly suitable, as these have strong action-sound links that take advantage of people's past knowledge and are more likely to be understood by a larger number of people. This is supported by the results of GSC2 and GSC3; as these types of gestures were validated in both studies.

9 - Conclusion

This thesis reviewed the relevant literature, specifically considering:

- Methods to identify common DJ techniques.
- DJ-audience communication.
- The use of gestures in music performance and production including existing gesturally controlled DJ systems.
- Gesture design methodologies including user centred design and in particular the GES methodology.

A GES was conducted to create an end-user gesture set in an attempt to improve DJ-audience communication. Unfortunately, the majority of referents did not receive many high consensus gestures and agreement rates were low. The GES resulted in conflicting and inconsistent gestures for the referents which prevented an end-user gesture set from being directly produced.

Consequently, three Gesture Set Creation (GSC) experiments were carried out. GSC1 was conducted to try to create a conflict free gesture set. Test subjects were asked to map any of the elicited gestures to any of the audio referents. Unfortunately, GSC1 still contained directionally inconsistent gestures, reinforcing the original GES results. GSC2 was conducted to try to create an inconsistency free end-user gesture set. Test subjects were asked to select one of the three or four highest consensus gestures from the GES on a referent by referent basis. This experiment successfully removed the inconsistencies but the conflicting gestures for the EQ, volume and filter referents were still present. A final gesture set creation experiment, GSC3, was performed in order to create a conflict and inconsistency free gesture set. Test subjects were asked to subjectively rate gestures from gesture sets that were constructed on the basis of highest score in GSC2, strongest preference in GSC2 and by experts based on GSC2. Gesture sets for complex audio referents and simple audio referents were subjectively rated separately. The best combined gesture set consisted of strongest preference in GSC2 for complex audio processing referents and expert designed based on GSC2 for simple audio processing referents. This combined set successfully created a DJ gesture set that had no conflicts or inconsistencies.

Specific conclusions of this research are:

- The overarching aim of this project was to identify whether there is a common body language that can be used to communicate DJ sonic outcomes to an audience. The inconsistencies and conflicts for these experiments suggest that this is not the case for all audio processes.
- Results suggest that moving hands up and moving hands down is universally understood as increasing and decreasing intensity; people appear to apply this gesture to multiple audio processes.
- This project shows there is validity in an gesture elicitation approach for generating ideas for gesture to control audio processing.
- For complex audio referents a wide range of metaphorical gestures were proposed. When users were presented with a selection of these gestures there was a clear preference which suggests that although users generated different gestures, when presented with a choice of gestures there was a consensus.
- Performing a GES to improve DJ-audience communication using audience members and sonic referents produced a gesture set with conflicting and inconsistent gestures. Arguably the audience members should not be used because they lack the specialist

knowledge required to interpret the sonic referents. However, it could be appropriate to use audience members in an evaluation stage.

- When performing GES with related abstract referents they can't be randomly presented to the test subject because this denies them the opportunity to propose interrelated gestures.
- Users prefer high consensus end-user elicited gestures as opposed to randomly selected and low consensus gestures for the complex audio referents.
- When designing experiments that have to be completed unsupervised on the internet the most important factors are to keep the test short and simple. This recommendation is based on the high dropout rate experienced in GSC1. Also, presenting a large number of gesture videos to be mapped to a large number of audio referents imposes too much cognitive load on test subjects which potentially leads to a high attrition rate.
- A combination of subjective ratings and consensus should be used as a measure for suitability of gestures.

Adopting the GES for DJ-audience communication has afforded the opportunity to assess its suitability for this task. The following recommendations are based on this experience. When performing a GES for DJ-audience communication in the future, the elicitation should be treat as an idea generation phase. Following this, a subjective rating based evaluation should be performed that involves audience members and DJs. Finally, the most suitable gestures should be selected using consensus and the subjective ratings.

10 - References

Ahmed, A. Y., Benford, S., & Crabtree, A. (2012). Digging in the Crates: An Ethnographic Study of DJs' Work. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*, Austin, Texas, USA, 1805–1814. https://doi.org//10.1145/2207676.2208314

Ali, A. X., Morris, M. R., & Wobbrock , J. O. (2019). Crowdlicit: A System for Conducting Distributed End-User Elicitation and Identification Studies. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI 2019)*, Association for Computing Machinery, New York, NY, USA, 1-12. https://doi.org/10.1145/3290605.3300485

Anthony, B. (2017). Mixing As A Performance: Creative Approaches To The Popular Music Mix Process. *Journal on the Art of Record Production, 11*. https://www.arpjournal.com/asarpwp/mixing-as-a-performance-creative-approaches-to-the-popular-music-mix-process/

Apel, D. (2015*). Corey Soljan at Airbeat One 2015* [Photograph]. Wikipedia. https://en.wikipedia.org/wiki/Code_Black_(DJ)#/media/File:Airbeat_One_2015_Code_Black_Corey_Soljan_DJ_by_Denis_Apel-1027.jpg

Apple. (n.d.). *QuickTime Player User Guide*. https://support.apple.com/en-gb/guide/quicktime-player/welcome/mac

Ars Electronica. (2014). *Kelly Snook (UK)* [Photograph]. Flickr. https://www.flickr.com/photos/arselectronica/14968143997

ATLAS.ti. (2021). *What is ATLAS.ti*. https://atlasti.com/product/what-is-atlas-ti/

Balin, W., & Loviscach, J. (2011). Gestures to Operate DAW Software. *Journal of The Audio Engineering Society 130,* London, UK. http://www.aes.org/e-lib/browse.cfm?elib=15923 bantonblud. (2009, August 9). *ANDY C first mix @ Matter 07 08 200 9*[Video]. YouTube. https://www.youtube.com/watch?v=TFpnXZwv0NQ

Bazeley, P. (2018) Using Qualitative Data Analysis Software (QDAS) to Assist Data Analyses. In: Liamputtong P. (eds) *Handbook of Research Methods in Health Social Sciences*. Springer, Singapore. https://doi.org/10.1007/978-981-10-2779-6_108-1

Be-At Tv. (2017, December 28). Richie Hawtin in the sea [Video]. Facebook. https://www.facebook.com/watch/?v=10156050753017474

Beamish, T., Maclean, K., & Fels, S. (2004). Manipulating music: multimodal interaction for DJs. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*, Vienna, Austria, 327-334. https://doi.org/10.1145/985692.985734

Beatport. (n.d.). *Home*. https://www.beatport.com/

Bell, A., Hein, E., & Ratcliffe, J. (2015). Beyond skeuomorphism: The evolution of music production software user interface metaphors. *Journal on the Art of Record Production, 9*. https://www.arpjournal.com/asarpwp/beyond-skeuomorphism-the-evolution-of-music-production-software-user-interface-metaphors-2/

Bennet, B. (2020, June 16). *Where do DJs get their music from?*. https://www.djgym.co.uk/post/where-do-djs-get-their-music-from

Beringer, N. (2001). Evoking Gestures in SmartKom - Design of the Graphical User Interface. *Revised Papers from the International Gesture Workshop on Gesture and Sign Languages in Human-Computer Interaction (GW '01)*, 228–240. https://dl.acm.org/doi/10.5555/647592.728738

Berndt, A., Waloschek, S., & Hadjakos, A. (2016). Hand Gestures in Music Production. *Proceedings of the International Computer Music Conference 2016*. http://www.cemfi.de/wp-content/papercite-data/pdf/berndt-2016-hand-gestures.pdf

Bluerooms Studio Huddersfield. (2016). *LMP Profile*. https://selene.hud.ac.uk/sengbr/Joomla3/index.php/bluerooms-profiles/lmp-profile

Boiler Room. (2013a, November 5). *Hot Since 82 Boiler Room x Warehouse Project DJ Set*[Video]. YouTube. https://www.youtube.com/watch?v=byAbXUeNLgM

Boiler Room. (2013b, December 12). *Skream Boiler Room London DJ Set*[Video]. YouTube. https://www.youtube.com/watch?v=b4hp75IwhV8

Boiler Room. (n.d.). *Home*. https://boilerroom.tv/

Bokowiec, M. A. (2011). V!OCT (Ritual): An Interactive Vocal Work for Bodycoder System and 8 Channel Spatialization. *Proceedings of the 2011 New Interfaces for Musical Expression (NIME '11)*, Oslo, Norway. https://www.nime.org/proceedings/2011/nime2011_040.pdf

Brewster, B., & Broughton, F. (2006). *Last Night a DJ Saved My Life: The History of the Disc Jockey* (2nd ed.). Headline.

BrineWR71 (2018, September 20) *Because watching a guy standing next to his MP3 player might make you question how much you paid to be* [Comment on the online forum post *Why do so many djs pretend to do much more than they're actually doing. Ive seen all sorts of hot*]. Reddit. https://www.reddit.com/r/Beatmatch/comments/9hgggk/why_do_so_many_djs_pretend_to_do_much_more_than/

Brown, D., Nash, C., & Mitchell, T. (2018). Simple mappings, expressive movement: a qualitative investigation into the end-user mapping design of experienced mid-air musicians. *Digital Creativity, 29*(2-3), 129-148. https://doi.org/https://doi.org/10.1080/14626268.2018.1510841

Bryan, N. J., & Wang, G. (2011). Two Turntables and a Mobile Phone. *Proceedings of the 2011 New Interfaces for Musical Expression (NIME '11)*, Oslo, Norway. http://mcd.stanford.edu/publish/files/2011-nime-turntable.pdf

Cafaro, F., Lyons, L., & Antle, A. (2018). Framed Guessability: Improving the Discoverability of Gestures and Body Movements for Full-Body Interaction. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, Montreal, QC, Canada, 1–12. https://doi.org/10.1145/3173574.3174167

Caramiaux, B., Susini, P., Bianco, T., Bevilacqua, F., Houix, O., Schnell, N., & Misdariis, N. (2011). Gestural Embodiment of Environmental Sounds: an Experimental Study. *Proceedings of the 2011 New Interfaces for Musical Expression (NIME '11)*, Oslo, Norway. http://articles.ircam.fr/textes/Caramiaux11a/index.pdf

Chan, E., Seyed, T., Stuerzlinger, W., Yang, X., & Maurer, F. (2016). User Elicitation on Single-hand Microgestures. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, San Jose California USA, 3403–3414. https://doi.org/10.1145/2858036.2858589

Chang, H. (2009). *Autoethnography as Method (Developing Qualitative Inquiry)*. Routledge.
Charlieekelly. (2013). *File:NI Traktor Kontrol S4 MK2 Controller Top.jpg* [Photograph]. Wikimedia commons. https://commons.wikimedia.org/wiki/File:NI_Traktor_Kontrol_S4_MK2_Controller_Top.jpg

Chen, Z., Ma, X., Peng, Z., Zhou, Y., Yao, M., Ma, Z., Wang, C., Gao, Z., & Shen, M. (2018). User-Defined Gestures for Gestural Interaction: Extending from Hands to Other Body Parts. *International Journal of Human–Computer Interaction, 34*(3), 238-250. https://doi.org/10.1080/10447318.2017.1342943

Choi, E., Kwon, S., Lee, D., Lee, H., & Chung, M. K. (2012). Can User-Derived Gesture be Considered as the Best Gesture for a Command?: Focusing on the Commands for Smart Home System. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56*(1), 1253–1257. https://doi.org/10.1177/1071181312561222

Cliff D.C. (2006). hpDJ: An Automated DJ with Floorshow Feedback. In: O'Hara K., Brown B. (eds) *Consuming Music Together. Computer Supported Cooperative Work*, vol 35. Springer, Dordrecht.

Csikszentmihalyi, M. (2002). *Flow: The Classic Work on How to Achieve Happiness, with a new Introduction by the author*. Ebury Pr.

DanceTelevision. (2013, August 12). *KiNK (Live) | Cocoon at Amnesia | Ibiza* [Video]. YouTube. https://www.youtube.com/watch?v=tXHhrNQnW4Y

DanceTelevision. (n.d.). *Home*. https://www.dancetv.net/

Davidson, J. W. (2012). Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: Two distinctive case studies. *Psychology of Music, 40*(5), 595–633. https://doi.org/10.1177/0305735612449896

Dewey, C., Wakefield, J., & Tindall, M. (2018). MIDI Keyboard Defined DJ Performance System. *Proceedings of the 2018 Conference on New Interfaces for Musical Expression (NIME '18)*, Virginia Tech, Blacksburg, USA. https://www.nime.org/proceedings/2018/nime2018_paper0043.pdf

Dezeen. (2014, April 27). *Imogen Heap's Mi.Mu gloves will "change the way we make music"* [Video]. YouTube. https://www.youtube.com/watch?v=ci-yB6EgVW4

Dim, N. K., Silpasuwanchai, C., Sarcar, S., & Ren, X. (2016). Designing Mid-Air TV Gestures for Blind People Using User- and Choice-Based Elicitation Approaches. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*, Brisbane, QLD, Australia, 204–214. https://doi.org/10.1145/2901790.2901834

DJ Mag. (2020, September 12). *Armin Van Buuren DJ Set From The Top 100 DJs Virtual Festival 2020*[Video]. YouTube. https://www.youtube.com/watch?v=eYEhOe9ahwk

DJ Play It. (2019). *CDJ vs Controller: Understanding the Key Differences & Similarities.* https://djplayit.com/cdj-vs-controller/#:~:text=You%20are%20able%20to%20perform,start%20rumbling%20the%20dance%20floor.

djdadi (2018, September 20) *Probably because DJ's are the only ones who actually know they aren't doing something, and we're a small minority of* [Comment on the online forum post *Why do so many djs pretend to do much more than they're actually doing. Ive seen all sorts of hot*]. Reddit. https://www.reddit.com/r/Beatmatch/comments/9hgggk/why_do_so_many_djs_pretend_to_do_much_more_than/

DJOKO (2020). Washed Away (Original Mix) [Song]. On *WASHED AWAY EP*. PIV https://www.beatport.com/release/washed-away-ep/2744959

do_not_engage (2018, September 20) *It's called "performing". Some performers lipsync. Some performers use backing tracks. Some performers use pre-recorded material. Some performers - MANY* [Comment on the online forum post *Why do so many djs pretend to do much more than they're actually doing. Ive seen all sorts of hot*]. Reddit. https://www.reddit.com/r/Beatmatch/comments/9hgggk/why_do_so_many_djs_pretend_to_do_much_more_than/

Dobrian, C., & Koppelman, D. (2006). The 'E' in NIME: musical expression with new computer interfaces. *Proceedings of the 2006 Conference on New Interfaces for Musical Expression (NIME '06)*, Paris France, 277–282. https://dl.acm.org/doi/10.5555/1142215.1142283

Ellis, C., Adams, T., & Bochner, A. (2011). Autoethnography: An Overview. *Historical Social Research / Historische Sozialforschung*, 36(4 (138)), 273–290. https://doi.org/10.17169/fqs-12.1.1589

fbonito (2011, May 2) *DJ Hands…* [Online forum post]. DJTechTools Forum. https://forum.djtechtools.com/showthread.php?t=29471

Feldmeier, M. C. (2003). *Large Group Musical Interaction using Disposable Wireless Motion Sensors* [MSc Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/33547

Fels, S., Gadd, A., & Mulder, A. (2002). Mapping transparency through metaphor: Towards more expressive musical instruments. *Organised Sound, 7*(2), 109-126. https://doi.org/doi:10.1017/S1355771802002042

Fintoni, L. (2015, January 23). *Wheel It Up: History of the Rewind*. https://medium.com/cuepoint/wheel-it-up-history-of-the-rewind-21fdcff243d9

Fonseca, W.V. (2010). *Dj Tiesto* [Photograph]. Flickr. https://www.flickr.com/photos/wvfonseca/4365150295

G0LDI_L0CKS. (2018, September 20) *Why do so many djs pretend to do much more than they're actually doing. Ive seen all sorts of hot* [Online forum post]. Reddit. https://www.reddit.com/r/Beatmatch/comments/9hgggk/why_do_so_many_djs_pretend_to_do_much_more_than/

Galkin, E. W. (1988). *A History of Orchestral Conducting: In Theory and Practice*. Pendragon Press.

Gates, C., Subramanian, S., & Gutwin, C. (2006). DJs' perspectives on interaction and awareness in nightclubs. *Proceedings of the 6th conference on Designing Interactive systems*, University Park PA USA, 70-79. https://doi.org/10.1145/1142405.1142418

George, C. (2019, August 22). *Richie Hawtin wants to revive techno's original, intimate spirit*. https://www.documentjournal.com/2021/06/invitation-to-the-voyage/

Gizmo, (2009, October 14). *CDJ-900 - How Slip Mode works*. http://skratchworx.com/newspage.php?p=archive&fn_mode=comments&fn_id=1352

Golden, E. (2012, June 17). *The Age Of The EDM Rockstar*. https://djtechtools.com/2012/06/17/the-age-of-the-edm-rockstar/

Golden, E. (2013, January 14). *The DJ Jesus Pose: A Joke, Or A Secret Weapon?*. https://djtechtools.com/2013/01/14/the-dj-jesus-pose-a-joke-or-a-secret-dj-weapon/

Graña, Y. Z. (2020). *Performer-initiated and audience controlled interaction effects in live-streamed music events* [Thesis, KTH royal institute of technology]. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1553710&dswid=471

Hansen, K. F., & Alonso, M. (2008). More DJ techniques on the reactable. *Proceedings of the 2008 Conference on New Interfaces for Musical Expression (NIME '08)*, Genova, Italy. https://www.nime.org/proceedings/2008/nime2008_207.pdf

Hayafuchi, K., & Suzuki, K. (2008). MusicGlove: A Wearable Musical Controller for Massive Media Library. *Proceedings of the 2008 Conference on New Interfaces for Musical Expression (NIME '08)*, Genova, Italy. https://www.nime.org/proceedings/2008/nime2008_241.pdf

Hertz (2019). Recreate (Slam Rework) [Song]. On *ARCHIVE EDITS LP.* Soma Records https://www.beatport.com/release/archive-edits-lp/2748195

Hunt, A., & Kirk, R. (2000). Mapping Strategies for Musical Performance. *Trends in Gestural Control of Music*, 21, 231–258. https://www.researchgate.net/profile/Andy-Hunt-2/publication/243774325_Mapping_Strategies_for_Musical_Performance/links/0a85e5398c19ec37f7000000/Mapping-Strategies-for-Musical-Performance.pdf

Ivanovic, L. (2006). *Hands* [Photograph]. Flickr. https://www.flickr.com/photos/lukatoyboy/330366669/in/album-1640123/

Jääskeläinen, R. (2010). Think-aloud protocol. *Handbook of translation studies*, *1*, 371-374. https://www.researchgate.net/publication/326381184_Handbook_of_Translation_Studies_-_Vol_1

Jack, R. H., Harrison, J., Morreale, F., & McPherson, A. (2018). Democratising DMIs: the relationship of expertise and control intimacy. *Proceedings of the 2018 New Interfaces for Musical Expression (NIME '18)*, Virginia Tech, Blacksburg, USA, https://www.nime.org/proceedings/2018/nime2018_paper0039.pdf

Jensenius, A. R. (2007). *ACTION -- SOUND - Developing Methods and Tools to Study Music-Related Body Movement* [PhD Thesis, University of Oslo]. https://www.duo.uio.no/handle/10852/27149

Jensenius, A. R., Wanderley, M. M., Godøy, R. I., & Leman, M. (2010). Musical gestures: concepts and methods in research. In R . I. Godøy, & M. Leman (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 12-35). Taylor & Francis.

Jones, M (2019a, November 1*). Techno Mix 1st November 2019*[Video]. YouTube. https://www.youtube.com/watch?v=avxP14Wv0mE&t=395s

Jones, M (2019b, November 23). *🌴#138 Tech House Mix November 23rd 2019🌴*[Video]. YouTube. https://www.youtube.com/watch?v=MnPuW6JdZJ8&t=137s

Katspaugh, (n.d.). *Overview*. wavesrufer.js. https://wavesurfer-js.org/

Kavanaugh, P. R., & Anderson, T. L. (2008). Solidarity and Drug Use in The Electronic Dance Music Scene. *The Sociological Quarterly, 49*(1), 181-208. https://doi.org/10.1111/j.1533-8525.2007.00111.x

King, P. (2009). *DJ Schiller* [Photograph]. Flickr. https://www.flickr.com/photos/pkingdesign/3522200465

Krefeld, V., & Waisvisz, M. (1990). The Hand in the Web: An Interview with Michel Waisvisz. *Computer Music Journal, 14*(2), 28-33. https://doi.org/10.2307/3679709

Kumar, A. B., & Morrison, S. J. (2016). The Conductor As Visual Guide: Gesture and Perception of Musical Content. *Frontiers in Psychology, 7*(1), 1049-1060. https://doi.org/10.3389/fpsyg.2016.01049

Kurtenbach, G. & Hulteen, E. A. (1990). *The art of human-computer interface design*, in B. Laurel, ed., '*Gestures in Human-Computer Communication*', Reading, PA: Addison Wesley, 309–317.

Laursen, L. F., Goto, M., & Igarashi, T. (2014). A Multi-Touch DJ Interface with Remote Audience Feedback. *Proceedings of the 22nd ACM international conference on Multimedia (MM '14)*, Association for Computing Machinery, New York, NY, USA, 1225–1228. https://doi.org/10.1145/2647868.2655065

Leap Motion. (n.d.). *Leap Motion Controller*. ultraleap. https://www.ultraleap.com/product/leap-motion-controller/

Lech, M., & Kostek, B. (2013). Testing A Novel Gesture-Based Mixing Interface. *Journal of The Audio Engineering Society*, 61(5), 301-313. http://www.aes.org/e-lib/browse.cfm?elib=16822

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology (The MIT Press)* (1st ed.). The MIT Press.

Lens, A. (2017, March 22). *Amelie Lens vinyl only home session* [Video]. YouTube. https://www.youtube.com/watch?v=RcVnqFxWMDM

Long, C. A., Landay, J. A., Rowe, L. A., & Michiels, J. (2000). Visual Similarity of Pen Gestures. *CHI '00: Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, The Hague, Amsterdam, 360-367. https://doi.org/10.1145/332040.332458

Lopes, P. A., Ferreira, A., & Madeiras Pereira, J. A. (2011). Battle of the DJs: an HCI perspective of Traditional, Virtual, Hybrid and Multitouch DJing. *Proceedings of the 2011 Conference on New Interfaces for Musical Expression (NIME '11)*, Oslo, Norway. https://www.nime.org/proceedings/2011/nime2011_367.pdf

Lopes, P. A., Ferreira, A., & Pereira, J. M. (2010). Multitouch Interactive DJing Surface. *Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology (ACE '10)*, Taipei Taiwan, 28-31. https://doi.org/10.1145/1971630.1971639

Luck, G., & Toiviainen, P. (2006). Ensemble Musicians' Synchronization With Conductors' Gestures: An Automated Feature-Extraction Analysis. *Music Perception, 24*(2), 189–200. https://doi.org/10.1525/mp.2006.24.2.189

Madesen, C. K., Geringer, J. M., & Wagner, M. J. (2007). Context specificity in music perception of musicians. *Psychology of Music, 35*(3), 441-451. https://doi.org/10.1177/0305735607072655

Maguire, M. (2001). Context of Use within usability activities. *International Journal of Human-Computer Studies, 55*(4), 453-483. https://doi.org/10.1006/ijhc.2001.0486

Mariano Pina, M. R. (2015). Entrepreneurs of the nightlife: An analysis of the work of Party Promoters and DJs in the city of São Paulo. *Special Issue 1: Social boundaries and meanings of work in the 21st-century capitalism, 1*(1), 105 - 117. https://wuwr.pl/fsoc/article/download/5826/5504/

MAXQDA. (2021). *Discover MAXQDA*. https://www.maxqda.com/what-is-maxqda

McGarry, G., Tolmie, P., Benford, S., Greenhalgh, C., & Chamberlain, A. (2017). "They're all going out to something weird": Workflow, Legacy and Metadata in the Music Production Process. P*roceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17),* Portland, Oregon, USA, 995–1008. https://doi.org/10.1145/2998181.2998325

McGrath, S., Chamberlain , A., & Benford, S. (2016). Making Music Together: An Exploration of Amateur and Pro-Am Grime Music Production. *Proceedings of the Audio Mostly 2016 (AM '16)*, Norrköping, Sweden, 186–193. https://doi.org/10.1145/2986416.2986432

Miller, M. (2012). *The Complete Idiot's Guide to Conducting Music* (2nd ed.). DK Publishing.

Mimu. (2021). *Gloves*. mimugloves. https://mimugloves.com/

Mitchell, T. J., & Heap, I. (2011). SoundGrasp: A Gestural Interface for the Performance of Live Music. *Proceedings of the 2011 Conference on New Interfaces for Musical Expression (NIME '11)*, Oslo, Norway. http://www.nime2011.org/proceedings/papers/M13-Mitchell.pdf

Mitchell, T. J., Madgwick, S., & Heap, I. (2012). Musical interaction with hand posture and orientation: A toolbox of gestural control mechanisms. *Proceedings of the 2012 Conference on New Interfaces for Musical Expression (NIME '12)*, University of Michigan, Ann Arbor. http://www.eecs.umich.edu/nime2012/Proceedings/papers/272_Final_Manuscript.pdf

Molina, P., Haro, M., & Jordá, S. (2011). BeatJockey: A new tool for enhancing DJ skills. *Proceedings of the 2011 Conference on New Interfaces for Musical Expression (NIME '11)*, Oslo, Norway. http://mtg.upf.edu/system/files/publications/G20-Molina.pdf

Momeni, A., & Henry, C. (2006). Dynamic Independent Mapping Layers for Concurrent Control of Audio and Video Synthesis. *Computer Music Journal, 30*(1), 49-66. https://www.jstor.org/stable/3682026

Moore, R. F. (1998). The dysfunctions of MIDI. *Computer Music Journal, 12*(1), https://doi.org/doi.org/10.2307/3679834

Morris, M. R. (2012). Web on the wall: insights from a multimodal interaction elicitation study. Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces (ITS '12), Cambridge, Massachusetts, USA, 95–104. https://doi.org/10.1145/2396636.2396651

Morris, M. R., Danielescu, A., Drucker, S., Fisher, D., Lee, B., Schraefel, M. C., & Wobbrock, J. O. (2014). Reducing legacy bias in gesture elicitation studies. *Interactions, 21*(3), 40–45. https://doi.org/10.1145/2591689

Morris, M. R., Wobbrock, J. O., & Wilson, A. D. (2010). Understanding Users' Preferences for Surface Gestures. *Graphics Interface Conference 2010, Canadian Information Processing Society*, CAN, 216-268. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/09/gi_gesture_preferences.pdf

Muser, S. (2015). Gestures in Human-Computer-Interaction. Ludwig-Maximilians-University of Munich, Munich. http://www.medien.ifi.lmu.de/lehre/ss15/ps/papers/Muser-GesturesInHCI.pdf

Nacenta, M. A., Kamber, Y., Qiang, Y., & Kristensson, P. O. (2013). Memorability of pre-designed and user-defined gesture sets. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*, Association for Computing Machinery, New York, NY, USA, 1099–1108. https://doi.org/10.1145/2470654.2466142

Nielsen, J. (2010). *Mental Models*. Nielsen Norman Group. https://www.nngroup.com/articles/mental-models/

NIME. (2021). *The International Conference on New Interfaces for Musical Expression*. https://www.nime.org/

Norman, D. (2013). *The design of everyday things* (2nd ed.). MIT Press.

Norman, D. A. (2010). Natural user interfaces are not natural. *Interactions, 17*(3), 6-10. https://doi.org/10.1145/1744161.1744163

NVivo. (2021). *About*. https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/about/nvivo

O, Gurin. (2020). *Man Playing Black Sound Maker*[Photograph]. Pexels. https://www.pexels.com/photo/man-playing-black-sound-maker-3695476/

Paine, G. (2009). Towards unified design guidelines for new interfaces for musical expression. *Organised Sound, 14*(2), 142-155. https://doi.org/10.1017/S1355771809000259

Petrov, P. (2016). Ayahuasca (Original Mix) [Song]. On *BREATH EP.* 303Lovers https://www.beatport.com/release/breath-ep/1761813

Pioneer DJ Sounds. (2014a, December 15). *Paul Oakenfold - DJsounds Show 2014*[Video]. YouTube. https://www.youtube.com/watch?v=UeWCEmw888U

Pioneer DJ Sounds. (2014b, September 8). *DJ Sneak live 90s mix on vinyl - DJsounds Show 2014*[Video]. YouTube. https://www.youtube.com/watch?v=UdWu09bUjPw

Pioneer DJ Sounds. (2016, November 10). *DJsounds Show 2016 - ANNA - special vinyl only set!*[Video]. YouTube. https://www.youtube.com/watch?v=UvGwlK0OC_A

Pioneer DJ Sounds. (2017, March 22). *Huxley - DJsounds Show 2017* [Video]. YouTube. https://www.youtube.com/watch?v=741pZwJAUp0

Pioneer DJ Sounds. (2018, January 3). *Cristoph - DJsounds Show*[Video]. YouTube. https://www.youtube.com/watch?v=gHKX5ST7z9s

Pioneer DJ. (2014a, May 20). *Pioneer DJ History - Part 1: Evolution of the CDJ* [Video]. https://www.youtube.com/watch?v=CgTdAdynUYo

Pioneer DJ. (2014b, Nov 28). *Pioneer DJ History - Part 2: Cut the Mid Range, Drop the Bass* [Video]. https://www.youtube.com/watch?v=XO6kXAZdw5Q

Pioneer. (n.d.a). *XDJ-700*. https://www.pioneerdj.com/en-gb/product/player/xdj-700/black/overview/

Pioneer. (n.d.b). *DJM-800*. https://www.pioneerdj.com/en-gb/product/mixer/archive/djm-800/black/overview/

Piumsomboon, T., Clark, A., Billinghurst, M., & Cockburn, A. (2013). User-defined gestures for augmented reality. *Extended Abstracts on Human Factors in Computing Systems (CHI '13)*, Paris France, 955–960. https://doi.org/10.1145/2468356.2468527 polybius (2012, February 22) *Over exaggerating movements* [Online forum post]. DJTechTools Forum. https://forum.djtechtools.com/showthread.php?t=47233

Price, H. (2011). The Effect of Conductors on Ensemble Evaluations. *Bulletin of the Council for Research in Music Education*, (189), 57-72. doi:10.5406/bulcouresmusedu.189.0057

Ratcliffe, J. (2014). Hand Motion-Controlled Audio Mixing Interface. *Proceedings of the 2014 Conference on New Interfaces for Musical Expression (NIME '14)*, Goldsmiths, University of London, UK. https://www.nime.org/proceedings/2014/nime2014_518.pdf

Rovan, J. B., Wanderley, M. M., Dubnov, S., & Depalle, P. (1997). Instrumental Gestural Mapping Strategies as Expressivity Determinants in Computer Music Performance. *Proceedings of Kansei - the technology of emotion workshop*, Genoa, Italy, 68-73. https://www.researchgate.net/publication/2765549_Instrumental_Gestural_Mapping_Strategies_as_Expressivity_Determinants_in_Computer_Music_Performance

Ruiz, J., Li, Y., & Lank, E. (2011). User-Defined Motion Gestures for Mobile Interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, Vancouver, BC, Canada, 197–206. https://doi.org/10.1145/1978942.1978971

Sandor, C., & Nakamura, H. (2018). SoliScratch: A Radar Interface for Scratch DJs. *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, Munich, Germany, 427-427. https://doi.org/10.1109/ISMAR-Adjunct.2018.00129

sarasin (2011, December 20) *People who over-exaggerate and "flutter" or "flick" their hands while DJ'ing* [Online forum post]. DJTechTools Forum. https://forum.djtechtools.com/showthread.php?t=42839&page=5

Slayden, A., Spasojevic, M., Hans, M., & Smith, M. (2005). The DJammer: "Air-Scratching" and Freeing the DJ to Join the Party. *Conference on Human Factors in Computing Systems (CHI 2005)*, Portland, OR, USA , 1789–1792. https://doi.org/10.1145/1056808.1057023

Statista. (2018). *Share of households with a computer at home worldwide from 2005 to 2019*. https://www.statista.com/statistics/748551/worldwide-households-with-computer/

Statista. (2021). *Digital DJ controller wholesale sales in the United States from 2010 to 2020 (in million U.S. dollars)*. https://www.statista.com/statistics/443669/digital-dj-controllers-us-wholesale-sales/

Sturman, D. J., & Zeltzer, D. (1994). A Survey of Glove-based Input. *IEEE Computer Graphics and Applications, 14*(1), 30-39. https://doi.org/10.1109/38.250916

Theremin, L.S., & Petrishev, O. (1996). The Design of a Musical Instrument Based on Cathode Relays. *Leonardo Music Journal* 6, 4950.
https://www.muse.jhu.edu/article/585354

Thomas, D. R. (2003). *A general inductive approach for qualitative data analysis*.

Thompson, M. R., & Luck, G. (2012). Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. *Musicae Scientiae, 16*(1), 19–40. https://doi.org/10.1177/1029864911423457

Thompson, W. F., & Russo, F. A. (2007). Facing the music. *Psychological Science, 18*(9), 756–757. https://doi.org/10.1111/j.1467-9280.2007.01973.x

Till, R. (2010). *Pop Cult: Religion and Popular Music* (Illustrated ed.). Continuum.

Tindall, M. (2020, November 6). *Require help from fellow music technologists*. https://www.reddit.com/r/Techno/comments/jp7x6x/require_help_from_fellow_music_technologists/

Tone.js. (n.d.). *Home*. https://tonejs.github.io/

Traxsource. (n.d.). *Home*. https://www.traxsource.com/

Vatavu, R. D., & Wobbrock, J. O. (2015). Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*, Seoul Republic of Korea, 1325–1334. https://doi.org/10.1145/2702123.2702223

Vatavu, R. D., & Zaiti, I. A. (2014). Leap gestures for TV: insights from an elicitation study. *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video (TVX '14)*, Newcastle Upon Tyne, UK, 131–138. https://doi.org/10.1145/2602299.2602316

Villar, N., Gellersen, H., Jervis, M., & Lang, A. (2007). The ColorDex DJ System: A New Interface for Live Music Mixing. *Proceedings of the 2007 Conference on New Interfaces for Musical Expression (NIME '07)*, New York, NY, USA. https://doi.org/10.1145/1279740.1279795

Villarreal-Narvaez, S., Vanderdonckt, J., Vatavu, R. D., & Wobbrock, J. O. (2020). A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies?. *Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20)*, Eindhoven Netherlands, 855–872. https://doi.org/10.1145/3357236.3395511

Vogiatzidakis, P., & Koutsabasis, P. (2019). Frame-Based Elicitation of Mid-Air Gestures for a Smart Home Device Ecosystem. *Informatics, 6*(2), 23-39. MDPI AG. Retrieved from http://dx.doi.org/10.3390/informatics6020023

Waisvisz, M. (1985). The Hands: A Set of Remote MIDI-Controllers. *ICMC*.

Wakefield, J., Dewey, C., & Gale, W. (2017). LAMI: A Gesturally Controlled Three-Dimensional Stage Leap (Motion-Based) Audio Mixing Interface. Journal of The Audio Engineering Society 142, Berlin, Germany. http://www.aes.org/elib/browse.cfm?elib=18661

Walton, T., & Evans, M. (2018). The role of human influence factors on overall listening experience. *Quality and User Experience, 6*(1), 1-16. https://doi.org/10.1007/s41233-017-0015-4

Webber, S. (2018). *DJ Skills: The essential guide to Mixing and Scratching* (1st ed.). Routledge.

Wilson, J., & Bromwich, M. (2000). Lifting Bodies: Interactive dance – finding new methodologies in the motifs prompted by new technology – a critique and progress report with particular reference to the Bodycoder System. Organised Sound, 5(1), 9-16. doi:10.1017/S1355771800001035

Wobbrock, J. O., Aung, H. H., Rothrock, B., & Myers, B. A. (2005). Maximizing the guessability of symbolic input. *Extended Abstracts on Human Factors in Computing Systems (CHI '05)*, Portland, OR, USA, 1869–1872. https://doi.org/10.1145/1056808.1057043

Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009). User-Defined Gestures for Surface Computing. *Proceedings of the SIGCHI conference on human factors in computing systems*, Boston, MA, USA, 1083-1092. https://doi.org/10.1145/1518701.1518866

Wobbrock, JO. (2019). *About*. Crowdlicit. http://depts.washington.edu/acelab/proj/crowdlicit/

Wu, H., & Wang, J. (2013). Wu, H., & Wang, J. (2013). Design of Bare-hand Gestures for Object Manipulation in Virtual Reality ⋆. *The Journal of Information and Computational Science*, 10, 4157-4166. https://doi.org/10.12733/jics20102747

Wu, H., Zhang, S., Liu, J., Qiu, J., & Zhang, X. (2019). The Gesture Disagreement Problem in Free-hand Gesture Interaction. *International Journal of Human-Computer Interaction, 35*(12), 1102–1114. https://doi-org.libaccess.hud.ac.uk/10.1080/10447318.2018.1510607

YouTube. (n.d.). *Pioneer DJ Sounds - Home*. https://www.youtube.com/c/DJsounds/feature

Appendix A – GES Post Test Questionnaire

Name	

Age	

Gender	Male	Female	Binary	Prefer not to say

List your preferred music genres	

Describe your Djing experience (e.g. How many hours a week you practice, how many hours a week you DJ in a live environment).	

Describe your Experience as a music technologist (e.g. Do you produce/mix your own music? Do you own any equipment that you utilise at home)?	

Do you know what a HPF is?	
Do you know what a LPF is?	
Do you know what EQ is?	
Do you know what loops/looping is?	

How often do you go to a club/dance music venue? (0 been never and 5 been once a week or more)	0	1	2	3	4	5

	Yes	No

Do you feel there is any value in using gestures to control DJ performance?		

Would you be interested in using a gestural based system for Djing?	Yes	No
If yes what would you like it to control?		

Appendix B – Five Step Images of Highest Consensus Gestures

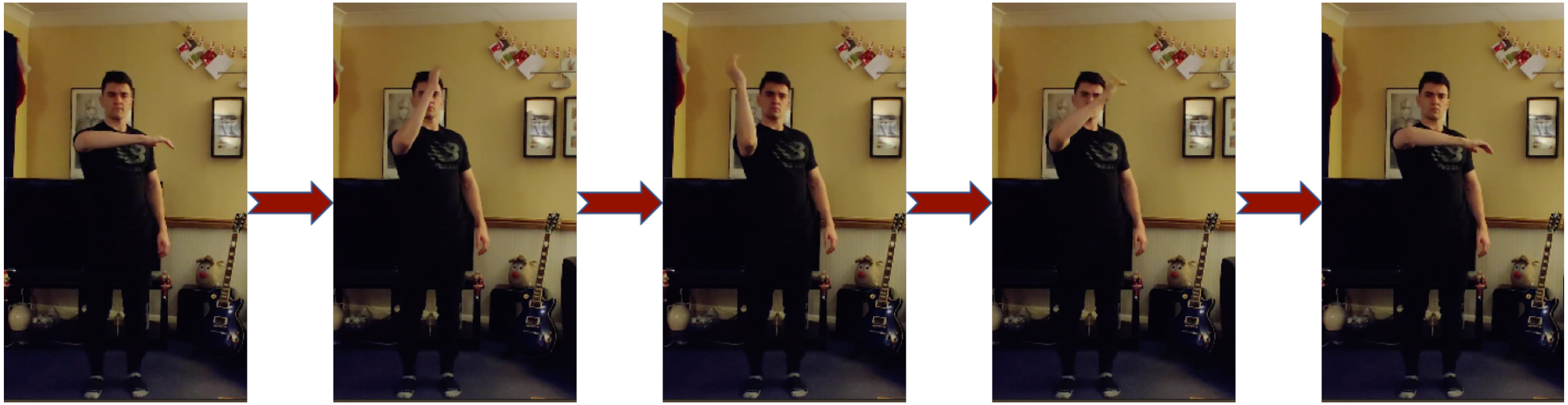


Figure B1. Five step images of Arm horizontal to arm vertical and back to horizontal (about elbow)

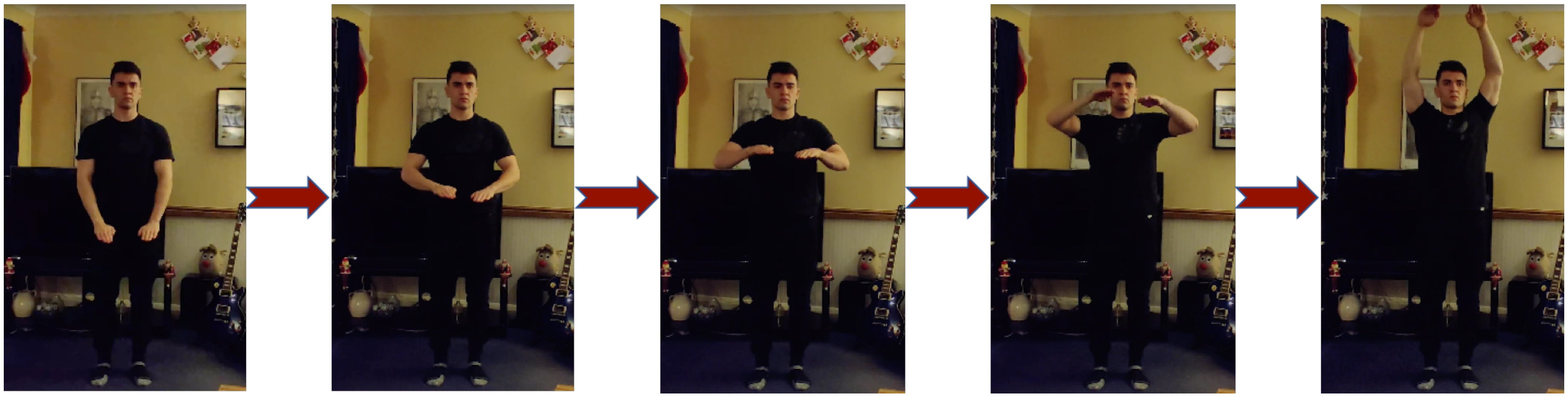


Figure B2. Both hands upwards

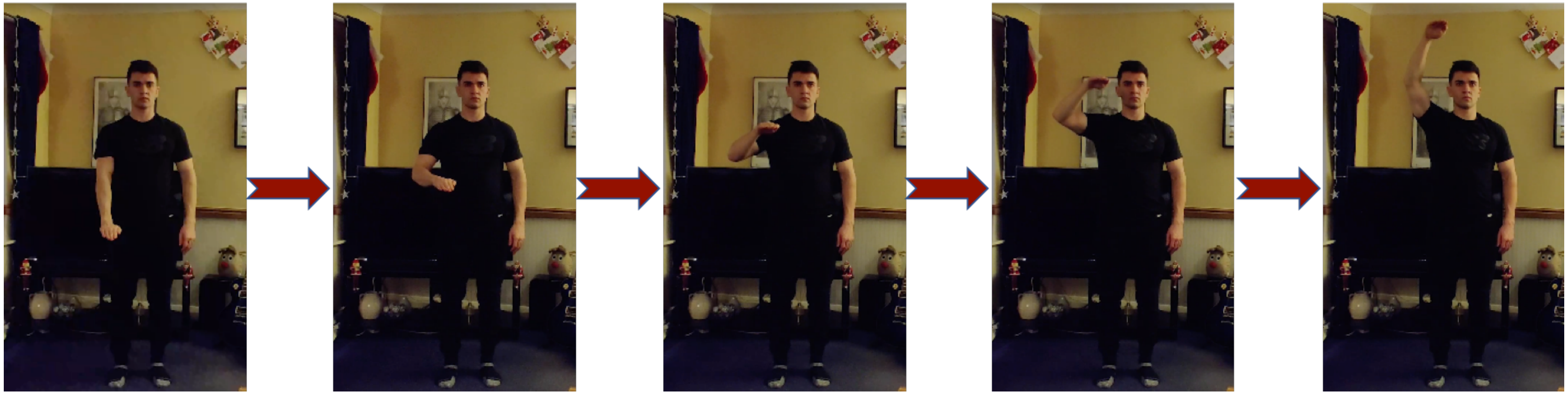


Figure B3. Single hand upwards

Figure B4. Both hands down in circular fashion

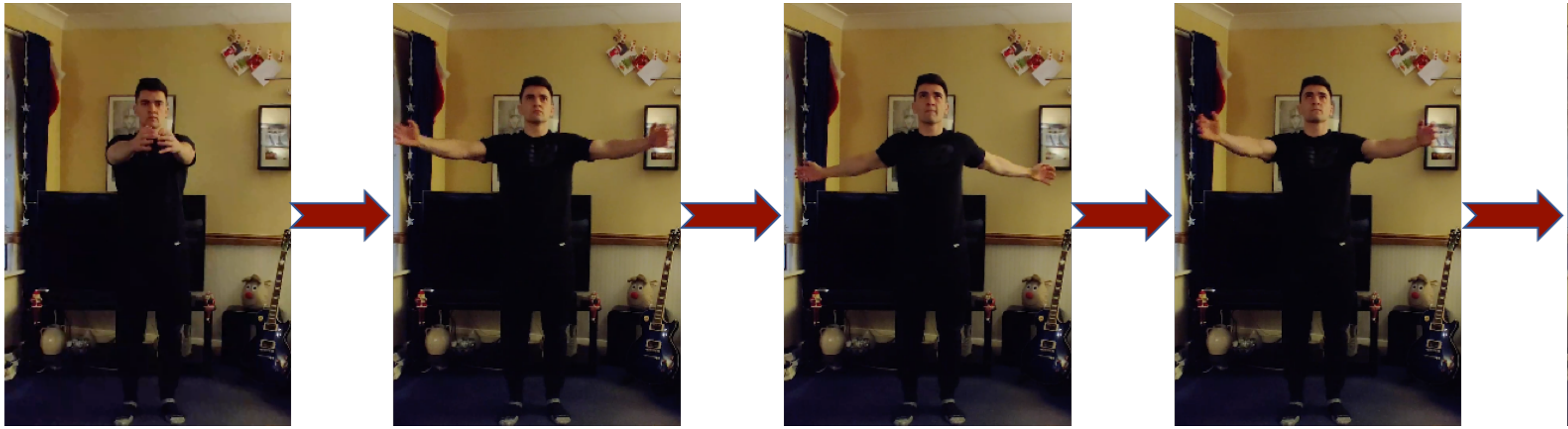


Figure B5. Hands centre to outstretched to side to centre again

Figure B6. Single hand downwards

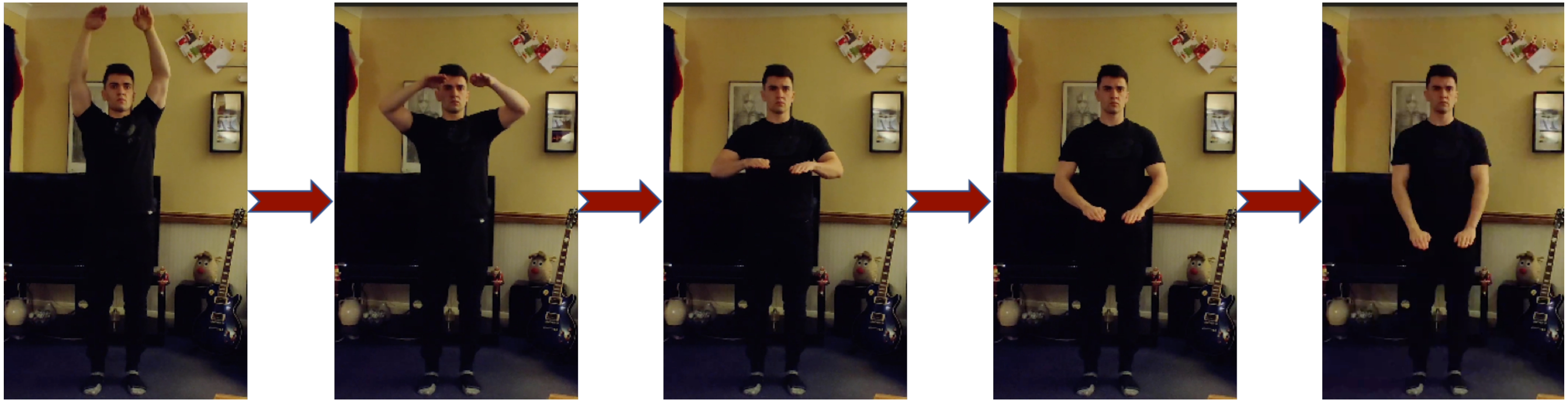


Figure B7. Both hands downwards

Figure B8. Body squats down

Figure B9. Knob turn metaphor

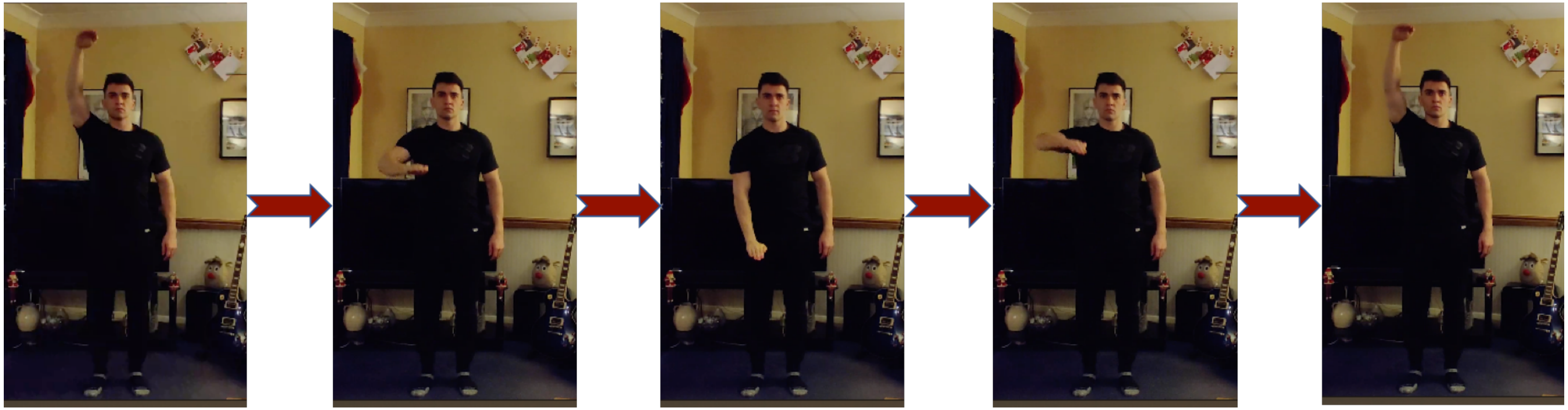


Figure B10. Single hand downwards then upwards

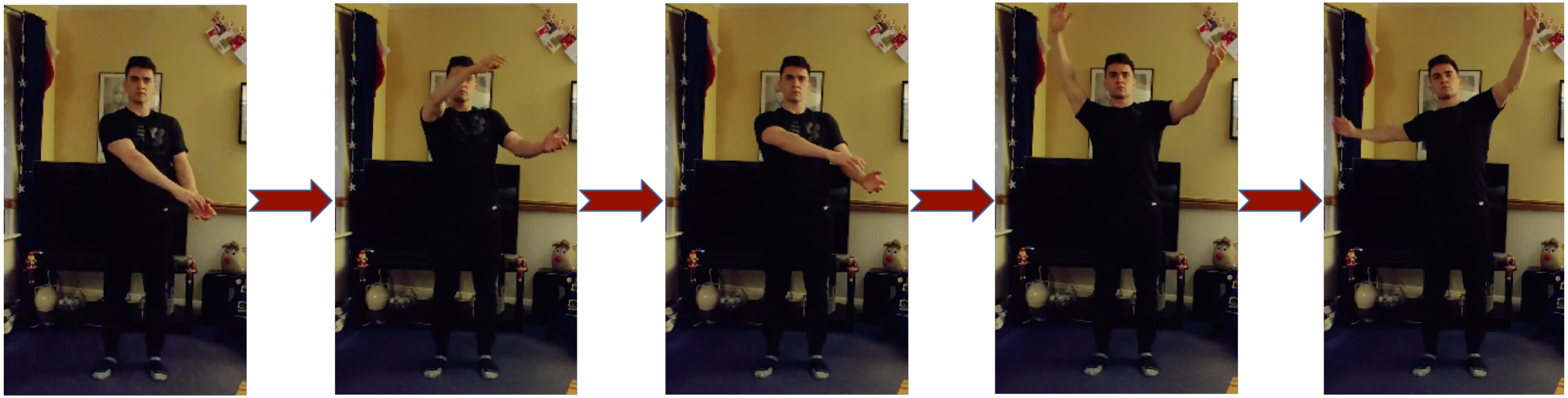


Figure B11. Little box big box whilst moving upwards

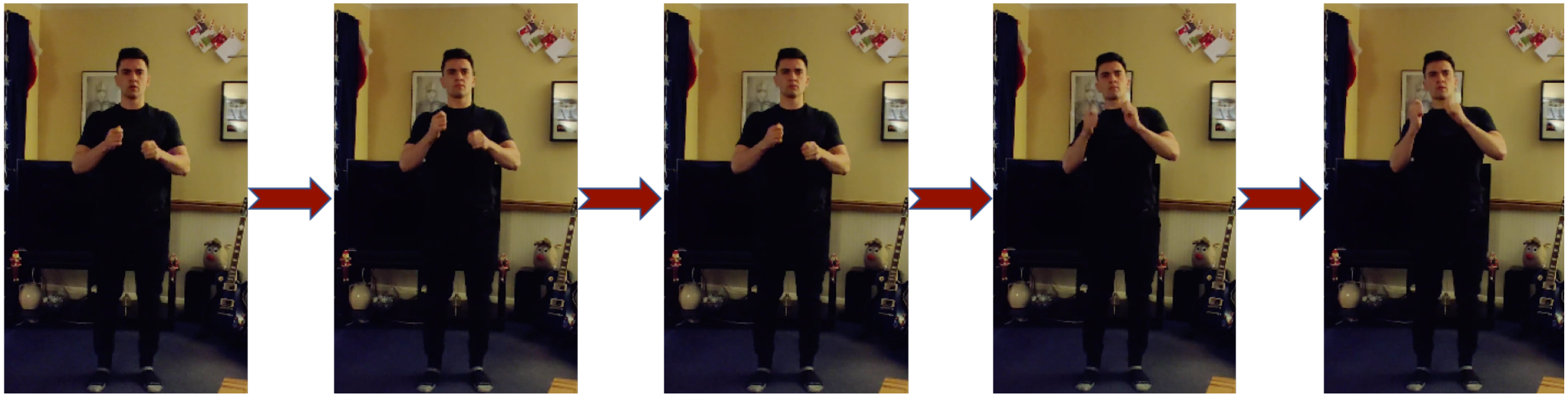


Figure B12. Drumming (as if holding sticks) metaphor

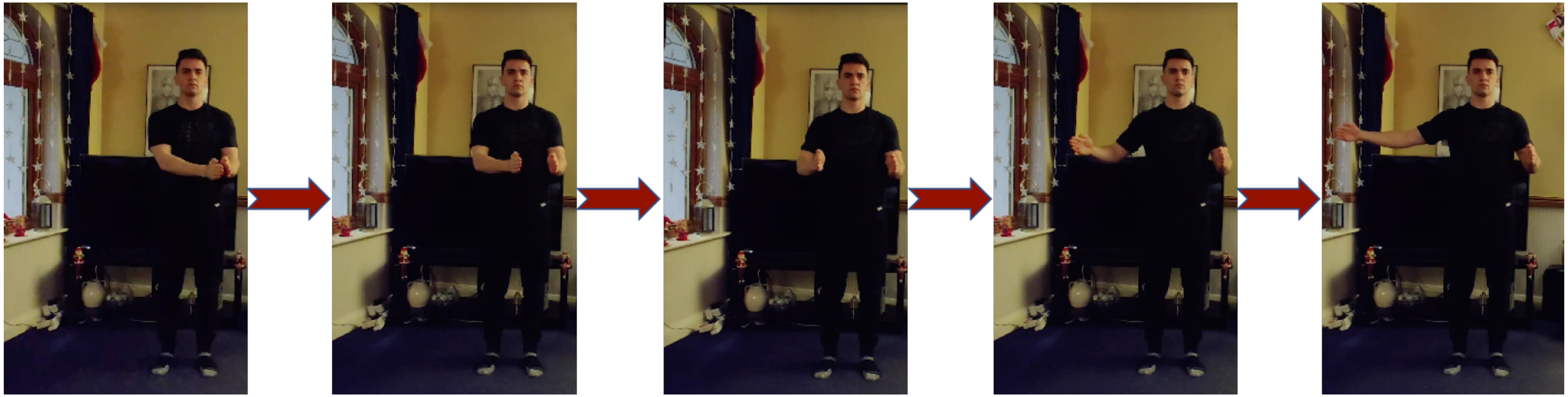


Figure B13. Little box, big box

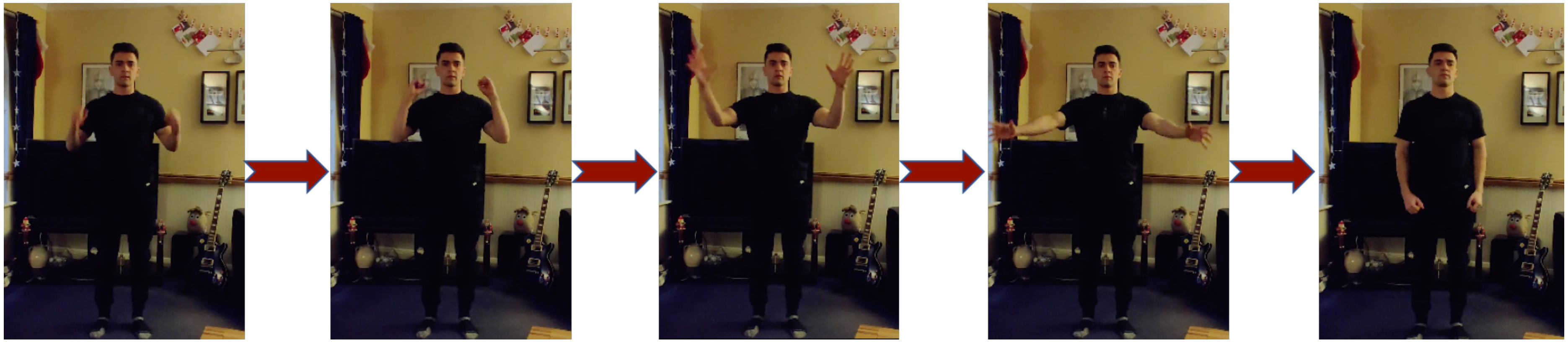


Figure B14. Both hands open and move out as if throwing something then close and move back in

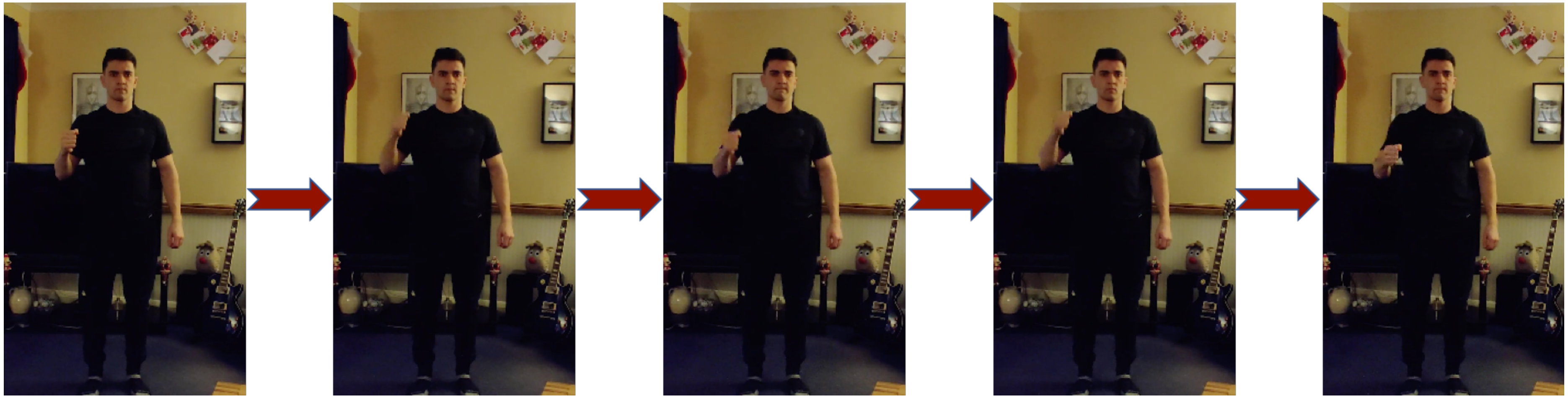


Figure B15. Shakes closed hand

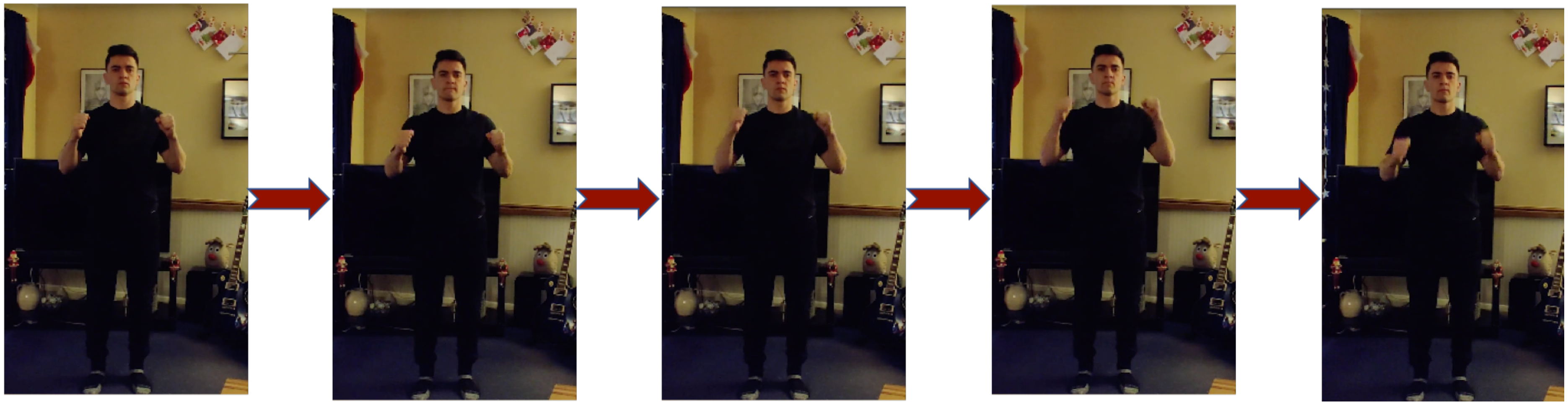


Figure B16. Shakes both hands

.

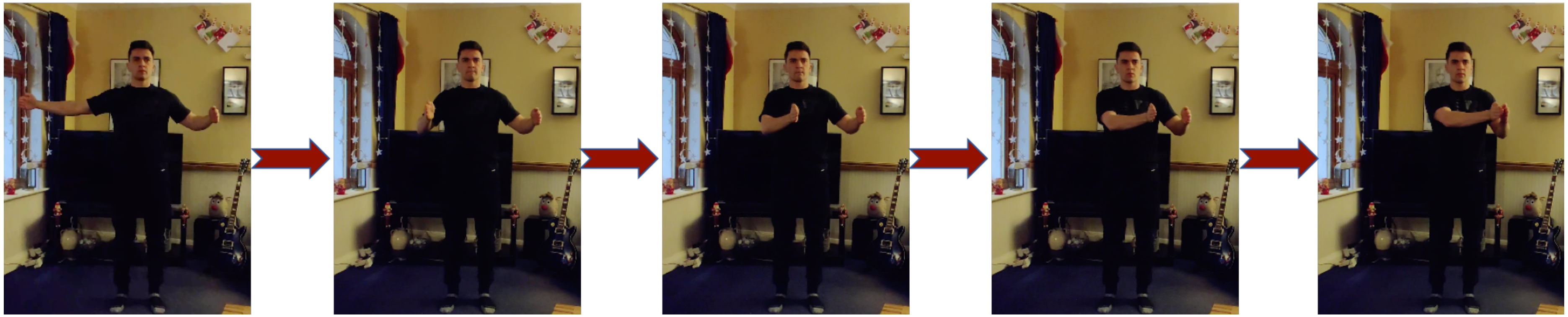


Figure B17. Big box, little box

Figure B18. Vinyl Spinback metaphor