Acoustic Instrument Simulation In Film Music Contexts

Boris Furduj

A thesis submitted in fulfilment of the requirements for the degree of Master of Research

Department of Media, Music, Communication and Cultural Studies, Faculty Of Arts, Macquarie University, Australia

October 2014

Table of Contents

List of tables

List of figures

Abstract

The purpose of this study is to investigate virtual acoustic instrument simulation in a film music context and to identify critical factors that impact the believability of music composed for the purposes of film scoring. To date, there has been no attempt to identify critical factors affecting 'believability' in the use of software-based virtual acoustic instruments, nor has there been any attempt to identify principles that might guide a composer's use of virtual instruments toward the goal of increasing the believability of virtual instruments in film contexts. The research aims to uncover what factors may influence the audience's recognition of real or virtual instruments, in order to call attention to methods that may be required for improving the believability of virtual instruments in digital film scoring in the future. The study implemented a mixed method approach consisting of practice-led research, aimed at gathering data on the use of composition production resources in the creation phase, and qualitative research methods, aimed at gathering data from listeners and viewers. Based on collected data and the broad range of possible variables tested, the study showed a number of elements that affected the perception of believability. As the music making process continues to shift and change, morphing into a continuous stream of activities that are enhanced by computer technology, this research will contribute to the diverse community of educators, researchers and practitioners who are working closely with new technologies in the fields of music and music technology education, music composition and the film industry.

KEYWORDS: Virtual Acoustic Instruments, Film Scoring, Believability

Declaration

I certify that the research presented in this thesis is original work carried out by the author. The work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university other Macquarie University and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

The research presented in this thesis was approved by Faculty of Arts Human Research Ethics Committee, reference number: 5201400530 on 28 May 2014.

Name: Boris Furduj Student ID: 42823919

Date: 10 October 2014

Acknowledgements

I am using this opportunity to express my gratitude to everyone who supported me throughout the course of this MPhil project.

This thesis would not have been possible without the help and patience of my principal supervisor, Professor Julian Knowles, who has supported me throughout my candidature and given constant motivation, support and advice, over and above what was required or expected. Julian’s editorial assistance in the preparation of this thesis is acknowledged gratefully.

Sincere thanks also go to my academic advisor, Professor Andrew Alter, for his help and guidance over the course of the study.

A special thank you to Jules for all your invaluable help, for being so approachable, for your kindness and golden advice.

Ana my wife, and my daughter Katarina, you took care of me and put your trust in me and made the journey seem easy and worthwhile. You are such an inspiration and a blessing to me. Lastly I’d like to thank my father Vladimir, my brother Fedja, and my other family in Belgrade, Branislav and Natasha.

Chapter 1: Introduction

> *"When image, dialogue, sound effects and music combine into multimodal texts, a 'chemical reaction' seems to take place. The resulting whole is, if maybe not greater, certainly different than the sum of the parts"* (Wingstedt, Brändström, & Berg, p. 18).

The impact of computer technology is all around us; from education, to health care, to communication, transport and entertainment. It is therefore not surprising it has had a resounding impact on the art of music in the late twentieth and early twenty-first centuries. It has transformed how music is practiced, transmitted, preserved, and heard. Less and less frequently do we hear musical sound that has not, at some level, been shaped by technology. As computers improve and become faster and more powerful, they are becoming as integral to a musician's work as any other instrument. As a result, we hear computer-simulated music everywhere: on the radio, TV, in live productions, in concert performances, and of course, at the cinemas. When it comes to producing film scores, computer simulations of full orchestral ensemble performances have developed in recent years to tremendous levels of realism. The practice of using digital samplers in film production began in the late 1970s and these are seen to be the precursors of modern 'virtual acoustic instruments'. In a short space of time, entire studio orchestras have been replaced with digital instruments. While many argue that technology is now so advanced the average listener cannot tell the difference between real and virtual instruments, no systematic research has been attempted to understand whether this is in fact true. The question remains as to how believable virtual acoustic instruments are in the context of film music.

Believability is a term that measures the level of realism of a virtual environment (Kim at el, 2004). For the purposes of this study, that environment is the cinema. The idea of believability has long been studied and explored in literature, theatre, film, radio and other media, but not to the same extent in music. Given that the use of virtual acoustic instruments is so pervasive in modern music practice, it is important to understand how audiences experience these instruments and whether or not they perceive them to be ‘real’.

Our sense of sight is generally considered more reliable than our sense of hearing, after all “seeing has, in our culture, become synonymous with understanding” (Kress & Van Leeuwen quoted in Wingstedt et al, p. 18). This adds a layer of complexity when analysing the believability of film music created using virtual instruments, as the impact of visual elements on the audience should not be underestimated. For this study, the term ‘virtual instruments’ refers to software tools designed to emulate acoustic instruments. Unless otherwise specifically stated, this paper does not refer to algorithmic music systems, where both musical events and sounds are generated from algorithms but more specifically to sample based virtual instruments that are played and/or sequenced by human performers. When delimited as such, no existing academic literature addresses the question of the believability of virtual acoustic instruments. The existing literature only refers to algorithmic music systems, which lie outside the scope of this study.

The aim of this research project, therefore, is to conduct an initial investigation into virtual acoustic instrument simulation in film music contexts within the above context and to identify the critical factors, if any, that affect believability. The research aims to provide answers to the following questions:

- Does musical genre impact believability?
- Does the presence or absence of other visual or sonic elements impact believability?
- Do performance and production processes impact believability?
- Does the musical background of the audience have any impact on their ability to perceive production differences in the film music they hear?

The study implements a mixed method approach consisting of practice-led research, aimed at gaining data around the use of composition and production resources in the creation phase, and qualitative research methods aimed at gathering data from listeners/viewers. The capture of data from both research methods allows for comparisons to be made between creation and reception contexts, and to test the thresholds of believability from 'expert' composer and 'non-expert' listener standpoints. These findings will commence the process of addressing the current absence of literature in an increasingly important area of creative and professional practice. Insights into the research questions are gained both from a practitioner driven investigation that encompasses the composition and production phases and from systematic listening tests conducted with typical cinema attendees.

The paper has been divided into six chapters. Chapter 2 provides a context review of the history of ideas that surround acoustic musical instrument simulation and maps the major thoughts that have emerged from both practice and technological developments. It also contributes a review of relevant academic literature supporting these ideas and developments. Chapter 3 describes the research design of the study, explaining the research methods employed, the rationale for the selection of participants, and the data collection procedures employed. Chapter 4 provides insights from the practitioner’s perspective on the composition of the cues used for the study, while Chapter 5 discusses the results and the findings. Chapter 6 offers some concluding commentary and suggestions for further research.

Chapter 2: Literature and context review

This chapter will be in two parts. The first part will provide a context review of the history of ideas that surround acoustic musical instrument simulation and will map the major thoughts that have emerged from both practice and technological developments. The second part will be a review of relevant academic literature supporting these ideas and developments. As will be seen, a review of the prevailing academic literature reveals there is a distinct lack of research that delves into the intricacies of creating believable film scores using computer generated acoustic instruments. As a consequence, this chapter will draw upon academic studies, professional sources and composer interviews as a way of establishing the prevailing views on the topic in the field.

2.1 Context review

Musical performance created entirely with computer-based instruments has been commonplace for the past 30 years. This is a result of the availability of inexpensive computing hardware and new software for real-time sound synthesis and manipulation (Cook, 2001). Today, these tools are so refined they can be used by professionals to create, record and edit audio. Some argue that with advances in algorithmic systems, computers can essentially “play and produce themselves” (Ramshaw, 2006). New generations of musicians and composers who might never have experienced a human to human music interaction are making music by interacting with technology only. Despite this, vast areas of music technology and listener experience remain completely uncharted from an academic perspective. Although there has been some interest in the area, there has been little sustained scholarship.

The notion of simulating sounds, musical instruments, or human performers is not a new phenomenon. During Roman times, one of the oldest instruments, the *hydraulis*, was an early pneumatically powered pipe organ. Modifications of this organ replicated birdsongs, and in the early nineteenth century mechanical organs called *orchestrions* imitated all the instruments of the symphony orchestra (Davies, 1996, p. 4). Similarly, instruments that can replicate voice and song have been around for centuries. In ancient China, a myth abounds of a bamboo spike fixed below a sliding temple door that was made to run in a groove in the floor, which produced the sounds 'please close the door' and, in reverse, 'thank you for closing the door' (Davies, 1996, p.5). In the late nineteenth century, Thomas Edison invented the cylinder phonograph, which is considered to be the first system ever devised for both storing and replaying sound, and as such, an important landmark technology in the effort to simulate or 'reproduce' human musical performances. Of course, one of the most widely utilised forms of simulating another instrument was developed during the second half of the twentieth century, when imitations of earlier instruments became widely available on electronic keyboard instruments.

Technological advancements are not reserved for popular music forms, such as techno, dance music, house or hip hop. There is a long history of technological innovation in the classical music tradition stretching back to the early twentieth century when developments such as *musique concrete* and *elektronische musik* changed the music landscape irreversibly. Pierre Schaeffer is considered a pioneer of sample-based music compositions and one of the most influential experimental, electroacoustic and electronic musicians. In 1948 he coined the term *musique concrete*, meaning music produced from recordings of instrumental and field recorded sounds which are then altered using tape techniques (Russ, 2004, p. 9). Schaeffer was the first composer to make use of a number of editing techniques that gave birth to the idea of constructing musical works and ‘performances’ from a library of pre-recorded materials. He was also one of the first musicians to use recorded sound combined with other sounds to create a musical piece. Tape splicing and tape looping techniques were used consistently in his research, which were often referred as ‘sound collage’. Karlheinz Stockhausen and Luciano Berio are also considered early innovators of tape music compositions. They used recorded sound sources to produce *electronische musik*. This repertoire of works from the 1940s and 50s established the idea that the studio techniques could be used to create unique listening experiences that extended beyond the confines of a singular musical performance. According to McGuire and Pritts, “early electronic composers found that manipulation of recorded sound materials opened a whole new palette of sound sources for musical expression” (2008, p. 193). These studio compositions created imaginary sound worlds that did not exist in the physical realm.

Glenn Herbert Gould, a celebrated Canadian classical pianist, was one of the first musicians to explore the idea of studio-based performance. For Gould, recording and broadcasting were not supplementary to the concert hall, but rather distinct art forms that “represented the future of music” (Bazzana, 2013). In the 1960s and 70s he produced scores of albums, steadily expanding his repertoire and developing a professional engineer’s command of recording techniques. One of his motivations for abandoning live performance at a relatively young age was to pursue the potential of the recording studio to construct performances that extended beyond the limits of any one recorded take and were constructed from many different takes. He was empowered to control every aspect of the final musical piece by selecting different parts from various takes to produce a ‘perfect’ performance. Gould likened his method to that of a film director, arguing that just as a viewer does not perceive that a two-hour film was made in two hours, nor should the act of listening to music be any different (Kingwell, 2009, p. 151).

Gould conducted a series of listening experiments whereby musicians, sound engineers, and non-experts were asked to listen to recordings and determine if and where splicing took place within a recording. Based on their relationship to music, different groups of people gave different answers, but no group gave accurate answers. Despite the fact his conclusions were by no means scientific, Gould made an observation: "The tape does lie, and nearly always gets away with it" (Kingwell, 2009, p. 158). This is an important historical precedent in the field of acoustic instrument simulation in that although the instrument itself was ‘real’ (in that Gould performed on a real piano) the performance was not. It was constructed from many different source recordings or takes, yet the audience believed it was one coherent performance. Gould’s work in the area demonstrated that audiences could perceive highly constructed and edited materials as a ‘real’ performance.

The next historical step toward the idea of the virtual instrument was the idea of the virtual acoustic instrument, which had its genesis in the sampler – a device which allowed the recording and replay of real acoustic instrument sounds and gestures. This concept can be seen to have originated in an early analogue tape based device: the Mellotron. In his book, Guide to MIDI Orchestration, Gilreath explains that the Mellotron, which generated orchestral sounds using pre-recorded strips of analogue tape, was the first 'sampler' available to the public (Gilreath, 2006, p. 521). The Mellotron was designed to operate so that when a key is pressed, a connected tape is pushed against a playback head, similar to a tape recorder. As long as the key stays depressed, the tape is drawn over the head, and a sound is played. When the key is released, a spring pulls the tape back to its original position (Awde et al., 2008, p. 17). Mellotron set an important technological precedent and was used in the 1970s by bands such as Genesis, the Moody Blues, and Yes. It was only after The Beatles used the technology, however, that Mellotron became popular and widely used by commercial studios (Goodwin, 1988).

By the late 1970s, digital technology was starting to emerge in the context of music production. This allowed the concept of the sampler to be further developed. The Fairlight computer musical instrument (CMI), a digital sampling synthesizer, was designed by Peter Vogel and Kim Ryrie in 1979. The Fairlight CMI used the Qasar M8 as its prototype, another synthesizer that had proved to be commercial unsuccessful. Fairlight introduced sampling to the world of commercial music production in the 1980s and was used on numerous hit singles and albums (Leete, 1999). Composer, Jan Hammer, was an early adopter of the technology, using the Fairlight to compose the soundtrack for the television drama series *Miami Vice*. Arguably, it was during this period that the concept of believability emerged; the notion that it should not be possible for the audience to tell the difference between acoustic instruments and digitally created sounds. The Fairlight CMI was originally marketed as a device that could replicate the sound of real instruments. Kim Ryrie, Fairlight founder, once said, "we wanted to digitally create sounds that were very similar to acoustic musical instruments, and that had the same amount of control as a player of an acoustic instrument has over his or her instrument” (Audio Media, 1996).

While the quality of the Fairlight's samples was unrefined compared with today's standards, it was nonetheless praised for its ability to mirror real instruments "perfectly". As Ryrie describes it: "the orchestra-in-a-box syndrome" (Audio Media, 1996).

The success of the Fairlight inspired other developments in sampling, most concerned with the goal of emulating or replacing real acoustic instruments. For example, New England Digital modified their digital synthesiser, the Synclavier, to perform sampling, and E-mu introduced a sampling keyboard called the Emulator, a device whose name spoke loudly about the commercial motivations and aims for the development of digital sampling technology. Ensoniq introduced the affordable Ensoniq Mirage in 1985, which made sampling available to the average musician for the first time.

Another important precedent work was the Hyperinstrument project. Started in 1986, the goal was to design expanded musical instruments using technology to give extra power and finesse to professional performers (Machover, N.D.). Guitars, keyboards, percussion, strings, even the conductor, were augmented using the project. Such hyperinstruments have since been used by some of the world's foremost performers, such as the Los Angeles Philharmonic, Peter Gabriel and Yo-Yo Ma. Since 1992, the focus of the hyperinstrument group has morphed and the emphasis is now on building interactive musical instruments for non-professional musicians, students and the general public. In the late 1980s, sampling became a standard feature of the electronic keyboard, not only as customised instruments but also as an additional method of generating more complex sounds (Davies, 1996). Since 1988, synthesizers, electronic organs and pianos have increasingly featured both synthesised and sampled sounds. These are sometimes kept as separate groups of waveforms and sometimes fused together. By 1991, 80% of synthesizers were based on sampling synthesis combinations, and as the computing power increased while costs decreased, the variation between the two was obscured.

Electronic sound synthesis incorporates many forms of sound production and sound processing methods, including subtractive synthesis, additive synthesis, wavetable synthesis, sample replay, and physical modelling. Physical modelling was developed to model the tonal characteristics produced by acoustic instruments, including all of their performance gestures. It was specifically targeted towards acoustic instrument simulation (Smith, 2004). Physical modelling synthesis can be regarded as modelling sound at its source, thereby enabling parsimonious representations and sonic manipulations that closely follow the physics of sound production. Consequently, physical modelling synthesis provides a complete playability range for virtual acoustic instruments" (Smith, 2004, p. 285). The first commercially available physical modelling synthesizer that modelled the sound of acoustic instruments through mathematical means was the Yamaha VL1 developed in 1994 by Yamaha in collaboration with Stanford University's Professor Julius O Smith III. It was the first synthesizer that modelled physical strings, wind and reed sounds. Many felt the VL1 produced sounds that so accurately reflected real instruments, it was hard to think of them as being electronically generated (Russ, 1994).

Another major advance in computer sample playback technology came when computer software company, NemeSyS, developed the computer software called GigaSampler in 1998. The GigaSampler was noteworthy because it didn't require a sample memory, instead directly streaming audio from the hard disk as required.

Prior to GigaSampler, hardware and software samplers required samples to be loaded in a random-access memory (RAM) device, which directly restricted the number and size of the samples they could use. The innovation of the GigaSampler meant loading times were greatly reduced, since rather than having to load sounds completely into RAM, the samples were read (streamed) from a hard drive and were always available. RAM was only used for temporary sample buffers. It was this product, according to Gilreath, that “included the streaming technology necessary to move sampling to the next level” (2006, p. 522). Streaming from hard disks enabled large sample collections to be ready for playback in real time.

In 1996, Steinberg GmbH, a German musical software and equipment company, released Virtual Studio Technology (VST), a software plugin standard that allowed the integration of software ‘instruments’ and effect plugins with digital audio workstation (DAW) software (e.g.,Cubase, Logic, Pro Tools and later GarageBand etc.) Software instruments became known as ‘virtual instruments’ allowed users to play and sequence sounds that were produced inside the DAW software itself. Virtual instruments could be used on the computer as plug-ins hosted by the DAW software or as stand-alone applications.

There are two kinds of virtual instruments. The first type generates sounds by creating and modulating waveforms – similar to traditional hardware-based synthesizers. The second type is sample-based, i.e., it triggers recorded audio samples, loops, and phrases performed by musicians. These samples are edited and assembled for use in a sample library, which can be accessed in real-time by the software instrument. In the case of an orchestral sample library, single notes and phrases are recorded with various expressions, tempos and articulations. These variations cover the capabilities of each instrument or ensemble. Finally, the recordings are edited in the studio and processed for use in a sample library or virtual instrument.

The demand for a closer integration between samples and software, driven by a desire for increased realism and usability, led to the development of a proprietary sample playback engines. These allowed composers, arrangers and music producers to reproduce a wide range of articulations and performance details employed by real soloists or ensembles. Following the stabilisation of the industry around a small number of different file standards and formats, these sample playback engines also allowed for the emergence of a wide range of sample libraries that would be produced by different software companies.

Different orchestral libraries have evolved over the years in such a way that they all exhibit different strengths and weaknesses for different applications, and what we find in practice is that composers will move around libraries to work to their strengths or to draw upon different libraries for different parts of the cue. No library is uniformly believable. Some patches and instruments are usually stronger than others and in to achieve increasing levels of believability, composers are utilizing different libraries for different applications and musical contexts, e.g. legato versus staccato, string quartet versus full orchestra or adagio versus allegro.

It is common for composers to deploy a range of different sample libraries, with choices influenced by the genre and the type of musical textures being created. Sometimes composers will create custom libraries that are adapted to their needs. For example, Hans Zimmer is well known for having created his own private collection of sampled orchestral instruments which gives him a signature sound because his samples are not used by anybody else. His personal library allows him to use different articulations and is not confined by the constraints of existing commercial libraries.

The current industry leaders of orchestral sample libraries include the Vienna Symphonic Library (VSL), LA Scoring Strings Collection, and the East-West Quantum Leap Symphonic Orchestral Library. Since its introduction in 2005, Vienna Symphonic Library (VSL) has been the most powerful virtual orchestral instrument on the market, consisting of a very large number of orchestral instruments and tens of thousands of articulations with a vast number of dynamic levels. The VSL library is recorded under studio conditions on the assumption that the sounds will be treated in post production according to the needs of the music cue and overall sound mix. The East-West Quantum Leap Symphonic Orchestral Library is recorded in well known orchestral halls and offers an acoustic, hall-based listening perspective without post production., while LA Scoring Strings Collection focuses on the inclusion of real time expression parameters and a comprehensive range of instrumental combinations. Each of these libraries therefore presents with different strengths depending on the application and context in which they are being used..

Nowadays, many classically trained composers are utilising virtual acoustic instruments (VAI) in their everyday work. One of the first examples of virtual musicians was presented in the Electric Garden at Siggraph 97 by the DIVA Group (Schertenleib, Gutiérrez, Vexo, & Thalmann, 2004). Since then, computer simulations of full orchestral performances have developed to high levels of realism thanks to advances in technology including, powerful audio/MIDI sequencers, detailed instrument reproduction through sound synthesis, and performance modelling using data acquisition and expressive performance rules (Sundstrup, 2009). Today, many composers use VAI as an illustration for what will ultimately be performed by a live ensemble. Composer Jerry Gerber, for example, argues that the expressive potential of VAI makes it an artistic medium in its own right, capable of creating a sound worthy of being the final composition (Wierzbicki, Platte, & Roust, 2012). Its precision, possibilities for new timbres and potential for fantastical automation make it a compelling platform for experimenting with and making music (Wang, 2007, p. 55). Ultimately, with the expanding potential of computers, orchestral simulations can be accomplished in real-time as believable virtual performances without requiring live instrumentalists' support.

Virtual orchestral instruments do not depend on physical constraints faced by their acoustic counterparts, such as membranes, strings or shape of the instrument. This fact permits a huge diversity of possibilities regarding sound production, but on the other hand strategies to design and perform these new instruments need to be devised in order to provide the same level of control subtlety available in acoustic instruments (Wanderley, 2001).

The development of virtual orchestral instruments is not confined to sample libraries. It extends to the corresponding technological development of MIDI and gestural controllers. When virtual instruments arrived, they were integrated into DAW environments and the input options were MIDI controller based (real time performance entry) or score/note entry based (step time entry). Therefore, the question of how to design and perform new computer based musical instruments consisting of gesturally controlled, real time computer generated sound needed to be considered in order to obtain similar levels of control subtlety as those available in acoustic instruments (Wanderlay, 2001). For example it has been possible since the 1980s to use gestural controllers modelled on wind instruments (e.g Yamaha WX-7) or similar alternate gestural controllers to capture breath articulation parameters to assist in producing more expressive wind and woodwind parts. Similarly a range of percussion controllers (e.g Roland Octopad) were developed to facilitate percussion based performance entry for drum and percussion parts.

Turning to cinema, the practice of using digital synthesizers in film production began in 1978 with the arrival of digital synthesizers and samplers (Palm, 2008, p. 63). Many issues, such as inadequate tuning and reliability, inherent in the analogue synthesizer were resolved and new possibilities opened up for film composers (Burt, 1995, p. 243). Entire studio orchestras were able to be replaced with this new technology. In 1982, Greek composer Vangelis, who used the synthesizer to replicate traditional instruments, became the first composer to win an Academy Award for an entirely digital soundtrack, *Chariots of Fire*. According to Hickman (2006, p. 382) synthesizers were widely used in 1984, with the scores for two of that year's top grossing films, *Beverly Hills Cop* and *Ghostbusters*, produced using the technology. The trend to use synthesizers and samplers in film scoring has continued in recent years. Sampled instruments and orchestral sample libraries have become an important part of composer and record producer Danny Elfman's sonic template. He used piano and drum sample libraries extensively on the soundtrack of 2009 film *Terminator Salvation*. Jeff Beal, one of the most prolific and respected composers working in Hollywood today, has utilized orchestral sample libraries for many years both in the mock-up and final phase of production. According to Beal, "these sounds offer absolute stunning quality, fidelity, and musicality" (Beal, 2014).

As this context review shows, the history of musical instrument and performance simulation has evolved significantly as a result of both practical and technological developments. Largely over the course of the twentieth century, successive refinements to key concepts and technology have enabled the evolution of an entirely new form of music production. In the realm of music performance, Glenn Gould demonstrated that performances could be convincingly simulated using editing techniques. These techniques became widely used in the analogue era and are now routine within the digital era. These developments highlighted that constructed/edited performances were readily accepted by audiences. At the instrument level, the concept of the 'sampler' emerged in the latter half of the 20th century and established that a musical instrument could be imitated by recording a library of notes, articulations and gestures. These would be re-assembled from firstly analogue tape technologies (Mellotron) and later via digital samplers (Fairlight, Synclavier, Emulator, Mirage) whose capacity became greatly extended via the emergence of 'streaming samplers' such as Gigasampler in 1997.

For sound synthesis, the industry saw a series of developments from modular analogue systems; synthesis pre-set systems, allowing the creation, storage and recall of 'imitation acoustic instrument' sounds; physical modelling synthesisers, which attempted to mathematically model the behaviour of real acoustic instruments; through to software synthesizers that allowed all synthesis approaches to be represented in software within a Digital Audio Workstation (DAW) system. While the invention of the synthesizer was not purely to imitate acoustic instruments, the uptake of synthesizers into commercial music production provided both the context and demand for them to be used to simulate acoustic instruments, thus offering a low cost, convenient way of using acoustic instrument sounds within orchestration without needing high budgets to pay for large numbers of musicians.

2.2 Literature review

This literature review aims to identify the key academic sources relevant to a study of believability in acoustic instrument simulation. Most of the available literature can be organised under five key themes: i) the advent and prevalence of computer generated instruments; ii) the notion of believability in virtual environments; iii) the inception of sampling and subsequent technological advances; iv) the precedent of believable virtual instruments; and v) the linkages between vision and hearing and their impact on an audience's power of perception. Ultimately, this chapter will reveal a significant gap in the literature that can be filled with this piece of research.

2.2.1 Believability

The idea of believability has long been studied and explored in literature, theatre, film, radio and other media (Bates, 1994, p. 1), but not music. Computer generated imagery has become an essential feature in the world of film screen content and interactive entertainment. In design fields, designers and architects strive to make 3D elements as realistic as possible in their computer assisted design drawings and animations (Pedersen, 2013). Although these elements may not be a real image of something we perceive from our own world, i.e. an alien or a dragon, we still accept them and because of that they become believable (Pedersen, 2013). The major goal of the virtual reality system is to simulate the sensory information such that the participant feels the generated experience was from the real world (Kim et al, 2004). As Bates (1994, p. 1) highlights, a "believable sequence does not necessitate an honest or reliable character, nor does it require human form, but one that provides the illusion of life and thus permits the audience's suspension of disbelief".

The same concept can be applied to music. According to Smith (2004) we appear to be approaching parity between real and virtual acoustic instruments in the context of recorded music playback. That is, we are approaching the time when many virtual instruments can be indistinguishable from their real-world counterparts for recording purposes under certain listening conditions (Smith, 2004, p. 45). Despite the increasing sophistication and prevalence of virtual instruments, there appears to have been no specific attempt to conduct research to identify the critical factors affecting 'believability' in the use of software-based virtual acoustic instruments. Nor has there been any attempt to identify principles that might guide the composer's use of virtual instruments toward the goal of increasing the believability of virtual instruments in screen contexts. The existing academic literature around software-based virtual instruments focuses almost entirely on experimental musical instruments, autonomous music systems, and performance interface design (McPherson & Kim, 2011; Collins, 2006). Furthermore, the literature does not directly address the core area of this investigation, which is acoustic instrument simulation in film music contexts and the key factors influencing believability.

The concept of believability in the context of acoustic instrument simulations only resides in literature in reference to spatial audio systems and the research surrounding 'virtual' positioning of acoustic sound sources in spaces defined by loudspeakers, headphones or a virtual reality system. In this sense, the term is used to refer to the success and accuracy with which a sound source may be positioned in a 'virtual' acoustic space (Pedersen & Jorgensen, 2005; Zhang & Johnston, 2013).

According to Savioja: “the goal is to deliver an acoustical message in a virtual reality system from the source to the receiver as it would happen in a real-world situation” (Savioja et al, 1999). This definition, unfortunately, does not extend to the synthesis or simulation of acoustic instruments per se, or the degree to which simulated instruments may be perceived by listeners as being ‘real’.

There are several sources that address algorithmic music systems (Sundstrup, 2009, Leonardo and Stefano, 2011) designed to model acoustic instrument behaviour but none which address sample based instruments in wide use by film composers.. What is also absent from the literature is any attempt to analyse the believability of virtual orchestral instruments, and the many intricacies of performance and sequencing that may affect their believability. Above all, there is no clear definition in the literature of ‘believability’ and so one will be offered for the purpose of this study.

2.2.2 Sampling and instrument simulation

In an attempt to analyse believability, some sources have attempted to chart technological developments and their impact on a composer’s ability to generate realistic simulation. Sundstrup, for example, examined several software applications and the techniques used to quantify the composite process required to generate an impressive, convincing, and realistic music performance through sample-based orchestral simulation (2009). Similarly, there are a number of publications that document the process of creating computer-based music, the techniques required and factors that impact the quality of the final product. *Sound Synthesis and Sampling*, for example, acts as a reference guide to the many techniques and approaches that are used in both commercial and research sound synthesizers (Russ, 2004). Russ defines sampling as a sound generating process that was established in the early twentieth century and that can reuse existing sound by processing it, or by producing sound mechanically or electronically. Although invented for telephony purposes, it was not until the invention of the transistor in the 1950s that it became practical to convert continuous audio signals into discrete digital samples using pulse code modulation (Russ, 2004, p.41). As McGuire and Pritts explain, “sampling is the process of recording a sound source one part at a time, each part of which is then imported into a sampler” (2008, p.1).

Several researchers have examined the impact digital samplers have had on simulation, with many suggesting computer software applications have made hardware-based digital samplers redundant as a result of their larger sample storage capabilities and superior performance. As Russ points out, “the 21st century has seen a wide adoption of software sample playback as an alternative to hardware: either as plug-ins to software MIDI and audio sequencers, or as stand-alone ‘sample’ sequencers” (2004, p.41). Wavetable, or sampling synthesis, is regarded as the most popular of the technology (Smith, 2004). According to Smith, the great advantage of sampling synthesis is static fidelity. Because the quality of the sound produced is limited only by the quality of the original recorded sound, and any sound an instrument makes can be recorded, he argues there is no fundamental lack of generality in sampling synthesis (Smith, 2004). He does acknowledge, however, that the two major drawbacks of the technology are that it consumes large quantities of memory, and can be very expensive when trying to achieve full playability (Smith, 2004). Massie (1998) adds that for many musicians it can simply be too much work to capture a complete range of playing conditions.

2.2.3 Application of technology to orchestras

Beyond the physical technology connected to sampling and instrument simulation, Cook (2001) reports that the musical interfaces that we construct are also influenced greatly by other factors. These include the type of music we like, the music we set out to make, the instruments we already know how to play, and the artists we choose to work with, as well as the available sensors, computers, networks, etc. He goes on to argue that, ultimately, the music we create and enable with our new instruments can be even more greatly influenced by our initial design decisions and techniques (Cook, 2001). Orchestration includes all uses of sound as a means of evoking texture. Composer Jerry Gerber suggests that "whether you use samples of acoustic instruments, complex synthesized textures, voice, recordings of live instruments, or sounds occurring in nature, there's an art and a craft to assembling them in a meaningful and expressive way. The principles are the same, whether you're dealing with a virtual orchestra or a real one" (Wierzbicki et al, 2012).

Composer Michael Prager qualifies the power of modern virtual instruments by explaining, "as fantastic as these new tools are, ultra-realistic orchestral tracks don't just flow out the minute you install them...there's a big difference between a track that sounds like a real orchestra and a cue that does its job in a film" (Wirzbicki, 2012, p. 250). He goes on to suggest that while there are many ways modern composers can learn how to get the most out of their digital orchestra, each mechanism requires a knowledge of traditional instruments and orchestration, combined with "a mastery of the technological resources available to you" (Wirzbicki, 2012, p. 250). Today, most Hollywood directors, when previewing a composer's film score, expect the mock-up orchestra (mostly containing orchestral sample libraries), to give an accurate simulation of a real orchestral performance (Geiger, 2007, pg.136). Steven Scott-

Smalley, arguably one of the most influential contemporary film orchestrators, explains that in his experience, once a track has been approved, most directors expect the real orchestra to sound like the mock-up orchestra (Geiger, 2007, p. 137). In 2003, a Broadway musician strike threatened the ability of many shows to perform. Musical producers made provisions to rehearse their shows with virtual orchestra accompaniment, in lieu of live musicians. *Chicago* co-producer, Barry Weissler said the result was “terrific” and that the virtual orchestra "looks, feels and sounds the same" for the audience (Phillips, 2003). Others were not so confidant. *La Boheme*’s musical director, Constantine Kitsopoulos, said he was glad the public never heard the virtual version as it was so “artistically compromising” (Phillips, 2003).

2.2.4 Precedent of believable virtual instruments

Whether or not audiences deem virtual instruments to be virtually compromising, there is certainly evidence that suggests humans can be readily fooled into believing they are the real thing. In recent years, several studies have been conducted in an attempt to understand the extent to which generative systems can produce music that is detectable to the human ear as having been produced by instruments and performers. In 1996, for example, Hall and Smith conducted a listening test with 180 participants, providing them with ten pairs of blues tunes; one computer generated and one human generated. The authors argued that “if the model successfully captures the structure of blues melodies, then listeners should have trouble distinguishing between human composed tunes and computer tunes” (Ariza, 2009). Their results showed that people were unable to reliably distinguish between the two production methods.

Similarly, Pearce and Wiggins used a collection of musical examples to train a genetic algorithm-based system. They then employed a discrimination test to determine if the output of the system was distinguishable from the training examples. They found that human listeners, whether experts or novices, could not objectively evaluate musical similarity. Additionally, they concluded that "there are absolutely no perceivable features" that differentiate the human and machine compositions, and that these features may include "such elusive notions as aesthetic quality or perceivable creativity" (Pearce & Wiggings, 2001, p. 25). More recently, in 2008, Collins conducted listener surveys to evaluate the output of Infno, a generative music system specialized for synth-pop and electronic dance music (Collins, 2008). Their results corresponded to those of Hall and Smith as "no statistically significant results…were found to distinguish perception of human and computer generation."

2.2.5 Link between sight and sound perception

What the above studies did not factor in, however, was the impact of sight on an audience's perception of sound. According to Pinch (2004), there is an emerging interdisciplinary area that studies the material production and consumption of music, sound, noise, and silence, and how these have changed throughout history and within different societies. He suggests that this field is known as "sound studies". Furthermore, Pinch argues that when performing sound studies, it is nearly impossible to escape from the visual (Pinch, 2004, pg.637). It is widely accepted that human beings make social judgments on the basis of both visual and auditory information (Tsay, 2013). For most people, sight is considered the most important factor when judging visual art forms such as film, while hearing is considered more important when it comes to sound, such as music. But this isn't necessarily correct. Particularly when the two forms are combined together, such as the case with film scores. When it comes to musical performance, it is consistently reported that sound is the most important source of information. Yet, a study by Tsay (2013) found that when judging musical performance, people relied primarily on visual information. She found that humans have a “natural, automatic, and nonconscious dependence on visual cues” when judging musical performance, despite sound being consciously valued as the core domain content. Although numerous studies have shown a general positive effect of the visual component on the experience and evaluation of music, it was a 2012 study by Platz and Reinhard which actually quantified the influence of the visual component. Their study concluded that there was an average medium effect size for visual components on the evaluation rankings such as liking, expressiveness and overall quality (Platz & Reinhard, 2012). Similarly, but in a converse vein, the book *Believable Virtual Environment: Sensory and Perceptual Believability* argues that in the case of film the audio is as, or even more, important than the video because “the surrounding sound defines the environment all around you” (Kim at al 2004, p.4).

This has significant implications when measuring the believability of virtual orchestration when combined with an image, such as a film. Just as the music will affect how we see things, the visuals will also determine how we hear the music (Wingstedt, Brändström, & Berg, 2010, p. 2). Importantly, music used in multimedia such as film, television and computer games is rapidly becoming one of the largest sources of modern musical experience (Wingstedt, Brändström, & Berg, 2010). Though typically experienced on an unconscious and unreflected level, Wingstedt, Brändström and Berg (2010) have found that this kind of music actively contributes narrative meaning in multimodal interplay with image, speech and sound effects. Besides the image we are also making sense, or trying to make sense, out of the intricate interplay with aural modes such as spoken language, sound effects and music (Wingstedt, Brändström, & Berg, 2010). And yet, the effect sound and music has on viewers is largely un-researched. Murch (1994) suggests this is paradoxical given the power of sound and the undeniable technical progress it has made in the last sixty-five years. He argues that, "whatever virtues sound brings to the film are largely perceived and appreciated by the audience in visual terms – the better the sound, the better the image" (Murch, 1994). Despite this, the literature on how film scoring impacts the audience's sense of believability is scarce. Murch (1994) postulates that it is the "cinematic inversion of the natural order" which has meant the analysis of sound in films has always been peculiarly elusive and problematical, if it was attempted at all.

2.2.6 Conclusion

The literature shows that developments in music technology have had a tangible impact on filmmaking. From a purely economic standpoint, it is more affordable to compose film scores digitally. From an artistic standpoint, the literature shows that developments in music technology have made it possible for composers to make use of a much wider palette of sounds and instruments when composing a scene. Composers have welcomed the opportunity to experiment and express with a plethora of sounds at their disposal. At the same time, this new technology has encouraged the next generation of musicians and filmmakers to compose and produce films, who might otherwise not have done so.

This chapter demonstrates that the existing literature does not adequately consider, if at all, the factors that impact the believability of film scores developed using computer generated acoustic instruments. What it does highlight, however, is that current research exists that covers concepts from different, yet related, fields, which provides some useful concepts that may be used to commence an investigation in a musical context. Aspects of the discussion surrounding the concept of believability in 3D animation and computer-generated images, where designers and architects strive to make computer-generated elements as realistic as possible, can be useful when considering film scoring, where composers are attempting to meet the same challenges in different media. The practices that permit audience members to believe the illusion and see the images as real, are the practices we need to map and apply when appropriate to film music. Similarly, the notion that images impact what we hear, and sounds alter what we see, has significant consequences for measuring the believability of music when combined with other sound effects and moving images.

Chapter 3: Methodology

This chapter will describe the research design of the study, explaining the research methods employed, the rationale for the selection of participants, and the data collection procedures employed.

3.1 Research methodology

It was envisaged that insights into the research questions would be best gained from both a composer/practitioner driven investigation and listening tests conducted with typical film viewers. This is due to the fact that the study is ultimately about the audience's perception of music cues and the way in which the perception of believability is or is not influenced by different composition and production approaches. To fully understand the nature of these relationships, it is necessary to gain control over the production of musical materials as well as the analysis phase.

Accordingly, the study implements a mixed method approach consisting of: i) *practice-led research*, aimed at gathering data on the use of composition production resources in the creation phase, and ii) *qualitative research methods* aimed at gathering data from listeners and viewers. The capturing of data from both research methods allows for analysis in both the creation and reception contexts, and to test and compare the thresholds of believability from both the creator and listener standpoints.

The *practice-led* component is aimed at gaining insights into matters of composition and production and their impact on the reception and believability of music cues. These insights can only be gained by the researcher actively engaging in the process of composition and production, documenting the process as research data, and making it available for analysis alongside data generated from the listening test phase.

Listening tests could have been conducted using pre-existing music cues from other screen composers and this study could rely simply on those materials. However, such an approach would not allow the enquiry to probe what specific factors (if any) in the composition and production of a music cue might impact believability. To gain an understanding of the impact on believability of different arrangement and production approaches it is necessary to make the composition and production process part of the research design. In so doing, the cues can be systematically organized and created to provide the basis for an examination of the extent to which instrumentation or production processes impact the perception of believability.

The *qualitative methods* used in the listening test consisted of discrimination testing (O'Mahony & Rousseau, 2003), open ended response questioning (Mack et al, 2005) and the collection of demographic data and information on musical skills and prior musical education.

The research has been designed to generate and test a broad range of possible variables in the production and reception stages. These variables have been created to test the extent to which they might impact the believability of a given cue. The following variables are used in the production stage:

1. Instrumentation (acoustic, virtual)

2. Performance and production factors (ensemble, solo multi-tracked performance, individual MIDI keyboard performance, mouse based note entry)
3. Genre/style

These variables are intended to test the extent to which genre, instrumentation and performance factors influence believability. The following variables are used in the reception stage:

1. Presence or absence of other sound elements (sound FX, dialogue etc)
2. Presence or absence of visual elements

These variables are intended to test the extent to which the presence of other sound or visual elements affected the perception of believability. The research also sought to capture and analyse the following range of audience factors:

1. Ability to play a musical instrument
2. Level of musical proficiency
3. Formal music education

These variables are intended to gauge whether a listener's musical background or skills play a role in the perception of believability.

3.1.1 Practice-led research

Practice-led research can be used to generate insights into the nature of practice itself or into the context in which practice occurs. It leads to new knowledge that will provide operational significance for the practice in question (Candy, 2006). For the purposes of this study, composing the film score permits an augmented understanding of the creative process. This would not be possible unless undertaking the process first hand, and will add valuable insights, ensuring a more holistic research project.

Although a relatively new approach, practice-led research is an extensively used method in creative arts contexts (Smith & Dean, 2009). Practice-led research began to emerge in the 1970s when ‘first generation’ pioneering artists and designers saw the potential for exploring and developing practice through the process and framework of formal research (Gray, 1996, p. 1). Previously, outputs from practice-led inquiries were seen as unreliable and too difficult to recreate since the research was individually based and concentrated around content, methods or processes that appeared specific to ‘one-time’ creative contexts (Silverstone, 1985, p. 203). Current consensus, however, supports practitioner-led research methodologies as a way of generating a more complete body of knowledge and understanding (Bell, 2008, p. 176; Milech, 2004, p. 7; McIntyre, 2006, p. 1). In this instance, the research aim is to generate insights as to where thresholds might lie for believability with respect to different approaches to film music composition and production. As the composer, the author has an intimate knowledge of exactly how each piece of music was created; there is no ambiguity about how a particular cue was made. In this way, accurate conclusions can be drawn in respect of this realm that would otherwise not be possible if using another composer’s materials.

A potential downside of practice-led research is that if the practitioner is also the researcher, tensions may arise. As articulated by Gray (1996, p. 7), these tensions can take the form of “subjectivity versus objectivity; internal versus external; doing versus thinking and writing; and intuition versus logic”. To avoid such tension, and to provide independent data on the research question, this study also employed a listening test and qualitative questionnaire, so that data obtained from listeners can provide independent, and thus defensible, perspectives on the believability of the film score.

3.1.2 Qualitative research

A qualitative research approach allows a researcher to identify issues from the perspective of study participants and understand the meanings and interpretations participants apply to behaviour, events or objects (Hennick, Hutter, & Bailey, 2010, p. 9). Additionally, “it seeks to understand a given research problem or topic from the perspectives of the local population it involves” (Mack et al, 2005, p. 1).

This study employed a qualitative questionnaire to gather listener responses to prepared cues. This consisted of multiple choice and open-ended questions about the perception of believability in the music cues presented, while also gathering select demographic data which might be correlated to listener responses to see if an individual’s musical background or experience had any bearing on their pattern of responses. The advantage of a qualitative questionnaire method over other quantitative testing methods is that it is more flexible and gives participants the freedom to respond in a more elaborate way and provide greater detail. As Mack et al (2005, p. 4) outline, open-ended questions have the ability to evoke responses that are: meaningful and culturally salient to the participant; unanticipated by the researcher; and rich and explanatory in nature.

3.1.2.1 LISTENING TEST

The design of the listening test was adapted from established listening test approaches in the field of audio and music perception, most notably the method of

Discrimination Testing (DT). Discrimination testing is a widely used listening test method, developed for a range of different research contexts. On the one hand it has been frequently used to test whether listeners are able to distinguish minor perceptual differences between playback hardware or file compression schemes. In this context the testing is highly technical in nature. Listening tests such as ABX, XY, (AB), ranking, paired comparisons and semantic differential are all variants of DT and provide a simple, intuitive means to determine if there is an audible difference between audio signals. They are considered the standard psychoacoustic test for this purpose (Boley & Lester, 2009, p.1).

On the other hand DT has also been used to test whether listeners can identify differences in musical **content** and how it might have been created. For example, it has previously been employed to determine the ability of listeners to distinguish between music composed by humans and those composed by computer generated instruments (Hall & Smith, 1996). One limitation of such a test is that when explicitly asking people to distinguish between human and computer generated music, listener expectations of human and machine performance may influence their judgment. According to Collins (2006), subjective biases are hard to remove and these types of tests, as much as anything else, “revealed much about the individual subjectivities of the participants”. This study has adopted a listening test as a deliberate manoeuvre to test variables at the reception stage of the process, rather than just the production stage, which relies entirely on the author’s own perspective (Ariza, 2009).

These previous studies demonstrated that DT was applicable and effective in contexts where listeners were being asked to distinguish between real and virtual inputs, albeit with a slightly different research focus. The research questions framed in this thesis sit between questions of technical production (sound source) and human agency (played live versus sequenced versus constructed out of real time via multi-track recording). They seek to probe further whether instruments are heard as ‘real’ and also whether they are ‘played’ by a human.

Designing the listening material and its relationship to the questionnaire is the most important factor when preparing to execute a listening test (Boley & Lester, 2009). Precisely defining the question to be answered is fundamental to ensuring sufficiently informative responses are generated and defensible conclusions can be made. In this regard, it is vital for the questions to be as clear as possible, while specific enough to yield focused responses that might lend themselves to meaningful comparison. Possibly one of the most important, yet frequently omitted, steps is to identify outside aspects, such as the listening environment and playback system that could lead to an incorrect conclusion (Boley & Lester, 2009). Literature suggests that appropriate design of listening tests and conditions is crucial to producing useable data (Bech & Zacharov, 2007). Boley and Lester suggest that a listening test should be carried out in a suitable room for the particular sounds, that loudspeakers should be appropriate for the test (Boley and Lester, 2009, p. 2).

3.1.2.2 SAMPLE

The aim was to survey a typical cross section of movie attendees. 36 participants took part in a listening test of the research, the target age range was 18-60 years with an even gender balance. Cultural and socio-economic differences were not seen as being major factors in the research and so did not inform the selection of participants. Inclusion in the research project was based on availability for testing and an overall desire to obtain a range of ages and a gender balance.

3.2 Data collection procedure

3.2.1 Creative practice inputs

Two stylistically different music cues were created. The first cue was an ambient, slow-paced piece, the second a faster-paced waltz. The purpose was to investigate the extent to which musical style, tempo, or complexity of articulation influenced the believability of the cue. Different musical styles require different playing techniques, therefore different samples and articulations of acoustic instruments had to be used. For example, long, sustained legato notes are widely considered by composers to be easier to imitate with virtual instruments,[1] so the author was interested to know if a cue constructed using such musical materials would be more believable. Collecting responses from listeners on two stylistically different cues would provide insight into whether these factors might impact believability.

Each cue was realized in four different versions as outlined in Table 1 below.

Version 1	Five piece acoustic ensemble with performers playing together in a same room
Version 2	Five piece acoustic ensemble, where each performer is multi tracked individually
Version 3	Virtual acoustic instruments using a computer as an instrument, each instrument performed individually from an input device or gestural controller
Version 4	Sequenced multi track recording created from event based editing

Table 1. Cue versions adopted for the study

For two reasons a conscious decision was made to create a recording using real musicians despite the study being an investigation into computer generated/simulated instruments. Firstly, providing the listeners with a recording of

[1] The conclusion that some playing techniques are easier to imitate than others was drawn by the author who, himself is a film composer with experience composing such tracks, and additionally from reading interviews and having discussions with other composers about different production techniques.

both real musicians and instruments, and computer generated/simulated instruments, allows for comparisons to be made as opposed to listeners simply determining whether what they are hearing sounds believable. Secondly, having a cue created with real musicians established a control version against which virtual cues could be compared when analysing listener responses.

Two different recording techniques were employed in the creation of the acoustic and virtual instrument cues. The purpose was to distinguish instrumental sound factors from performance factors and to explore the extent to which performance factors may impact notions of believability in addition to sonic factors. That is to say, regardless of the sound source itself, does the way in which a cue is recorded affect the audience's perception of human agency and believability?

In relation to the acoustic sources, the performers were first recorded playing together live in the same room. A second version was created where they were recorded individually, with each performer multi-tracked separately and combined later through mixing.

In relation to the electronic/virtual sources, each instrument was performed individually from an input device or gestural controller allowing for musical expression and nuance to be captured from a series of individual performances of the constituent parts in an arrangement. In the second, the notes for each part in the arrangement were manually entered out of real-time using a screen based computer interface. In this version there was no real-time human performance involved in the production of the cue whatsoever. While the virtual instruments playing the cue were the same, the performance factors used in the production of the cue were strongly contrasting.

Through organising the cues in this way, a standard set of performance factors (ensemble, multi-track, sequenced) could be tested across both virtual and acoustic instrument sources to determine their level of influence on the perception of believability in addition to (and in isolation from) the purely sonic factors associated with the instrument timbres themselves.

3.2.2 Listening test

The listening test was approximately one hour in duration. It was held in a university computer lab environment with high quality audio playback. The playback environment was adapted, as far as possible, to recreate a cinema listening experience so that participants had the impression they were in a movie theatre. Audio levels were calibrated to meet cinema audio playback loudness standards (referred to as Dolby Level), and 5.1 surround sound was implemented in accordance with Dolby standards (Kerins, 2011). Necessary precautions were followed to ensure that the audio playback equipment was functioning properly and that playback levels conformed to the published Dolby specification that governs playback levels in all commercial cinemas.

In order to investigate if the presence or absence of visuals and other sound elements (sound FX, dialogue etc.) affected the perception of believability in music cues, the first part of the test consisted of listening to recorded musical examples in the context of other sounds and pictures, while the second part of the listening test consisted of listening to the same musical examples in isolation from other sound and visual elements.

Examples were played in groups of 4. In total, 16 examples were played. The following shows the list and order of cues played to listeners.

Superman	
1	Virtual instruments: Performed individually
2	Acoustic ensemble: Recorded multi-track
3	Virtual instruments: Sequenced
4	Acoustic ensemble: Performing together
Waltz	
1	Virtual instruments: Sequenced
2	Acoustic ensemble: Performing together
3	Virtual instruments: Performed individually
4	Acoustic ensemble: Recorded multi-track

Table 2. Video files

Superman	
1	Acoustic ensemble: Recorded multi-track
2	Virtual instruments: Performing individually
3	Virtual instruments: Sequenced
4	Acoustic ensemble: Performed together
Waltz	
1	Acoustic ensemble: Recorded multi-track
2	Virtual instruments: Sequenced
3	Virtual instruments: Performed individually
4	Acoustic ensemble: Performing together

Table 3. Audio files

3.2.3 Questionnaire

During the listening test, after each example was played, the participants were asked to complete a questionnaire consisting of 3 questions, multiple choice questions and open-ended question that related to the examples played (Table 4, Questions 1, 2 and 3). After each set of four examples participants were asked an additional question (Table 4, Question 4). After all 16 examples were played the participants were asked additional 3 questions (Table 4, Questions 5, 6 and 7). The questionnaire is provided below in Table 4.

1. Did you feel you were listening to:		
A. Acoustic instruments	B. Virtual acoustic instruments	C. Mixture of the two
2. What made you think it was acoustic instruments, virtual acoustic instruments or mixture of those two?		
3. How strongly do you think this cue conveys an emotional feeling?		
A. It conveys weak emotion	B. It conveys strong emotion	C. Not sure
4. Which of the four audio samples did you like best?		
A. Example 1	B. Example 2	C. Example 3
D. Example 4	E. No opinion	
5. Do you play any musical instrument?		
A. Yes	B. No	
6. Have you received any formal musical instruction?		
A. Yes	B. No	
7. If yes, how would you describe your level of ability?		
A. Professional	B. Keen amateur	C. Hobbyist

Table 4. Questionnaire

The questions were phrased in such a way so as to be clear and non-specialist, thus avoiding any confusion from the listener or bias towards someone with specialist musical knowledge.

Question 1: This was a discrimination question to determine if the participant could feel a difference between recordings realised with real musicians versus virtual acoustic instruments. By analysing the number of correct answers, we were provided with a simple and accurate evaluation of listener aptitude for selecting the correct cue.

Question 2: This was an open-ended question that gave the participant the opportunity to further elaborate on what elements they felt made a cue more believable in their own words. It therefore provided an opportunity for a more in-depth and meaningful response regarding their perception of the cue.

Question 3: The purpose of this question was to find out if the examples had a functional equivalence to convey the same amount of emotion even if they were not believable and to ascertain whether or not this basic function requirement could be met or can this only be achieved by real musicians?

Question 4: This question was aimed at determining if the participant had a preferred sample and, if so, whether a pattern emerged for a preference towards real or virtual instruments.

Questions 5 – 7: These questions provided an opportunity to capture a focussed set of listener data around musical background and training to see if they affected the patterns of responses and in particular whether their musical background had any correlation with their ability to hear differences in the cues and their perception of believability.

3.2.4 Conclusion

This chapter described the research design of the study in order to demonstrate that the research methods employed are both adequate and suitable to answering the project's hypothesis. The practice-led component is geared towards gaining insights into the composition and production of film scores and how these might influence the reception and believability of film scores. The qualitative survey of listeners was employed to gain insights from audience members to ascertain whether a discernible difference between production methods could be found and the extent to which combining sound and image, as well a person's musical expertise, influences their perception of believability in the context of film music.

Chapter 4: Composer perspective

This chapter provides insight into the deliberate choices made by the composer in the design of this study. Production techniques, and their advantages and disadvantages, are discussed and the rationale behind why certain decisions were made is provided.

In the field of screen composition there exists a set of tacit understandings and conventions around the use of virtual instruments. These have emerged from composers' experiences using the tools in production contexts and by experiencing their strengths and weaknesses. Many composers speak about their attitudes towards, and practices around, the use of virtual instruments in professional literature and interviews (Asher, 2010; Folmann, 2004; Isham, 2010; Meehan, 1999; Stewart, 1999). The pattern that emerges is that there are implicit understandings around what virtual instruments can and cannot do well. For example, virtual instruments tend to be sample based and therefore the attack characteristic of each note is the same, unless very large sample libraries are used which randomise the sample used for each event or draw on a large library of sample layers that are mapped against dynamic levels (Meehan, 1999).

One of the fundamental components of any instrument's personal character consists of the initial transients of an articulated tone (Sundstrup, 2004). Besides the harmonic content - that determines an instrument's unique sound character, the prime influence on persuasive expressive performance simulation is note onset articulation - also known as note attack. "The term articulation is used to describe the amount of legato/staccato with which a note is being played. It is defined as the ratio between the note duration (i.e. sounding duration) and the IOI" (Friberg, Bresin, and Sundberg, 150). The IOI (inter-onset interval) referred to by Friberg, Bresin, and Sundberg concerns the time from one note's decay to the next note's attack. Accordingly, articulation can be expressed as either the onset/offset of separate tones, or the onset/offset of one tone in relation to another. Consequently, articulation refers to single notes, note repetition, and performance transitions as discussed later in this chapter (Sundstrup, 2004).

.

Before the present orchestral sample libraries were, the most common notated articulations such as staccato, pizzicato, marcato, spicato - needed to be simulated by changing each note's attack, decay, sustain, and release (ADSR) times using envelope generators, a multi-stage controller that allows the synthesizer to control over time the amplitude of a waveform" (Pejrolo, and DeRosa, 133). However, with the development of detailed orchestral sample libraries, each type of instrument articulation is sampled with all the natural transients included for each dynamic level.

According to Pejrolo and DeRosa "One of the biggest problems with a virtual MIDI orchestra is the fact that it is always perfectly in tune", This is the reason why it is advised to make a virtual ensemble sound a bit 'worse' than it could through a subtle use of the detune parameter" (Pejrolo, and DeRosa, 142). Commonly detuning methods such as detuning each string section and solo instrument by a small amount or random detuning between solo instruments for the duration of the score can have a profound affect on an orchestral simulation.

While there is little or no academic literature that independently formalises these understandings, there is little doubt from a practitioner perspective[2] that virtual instruments are developing technologies, which at this point in time, cannot fully replicate the level of realism and expression of an acoustic instrument. This is not to say that under certain controlled conditions and contexts a virtual instrument can't appear 'real' – it is the focus of this study to begin to understand the influencing factors in respect of this issue in a scholarly context. It is therefore part of the craft of being a film composer to make effective decisions around the use of virtual instruments and the design of musical material so they can become believable as required.

Given that these practices are prevalent in professional practice, this chapter documents, analyses, and discusses how the composer/researcher has organised materials within these limitations to achieve the highest levels of believability, prior to testing the results on listeners. It will also allow a comparison of different composition and production approaches with the listener responses outlined in the following chapter to see the extent to which these might influence believability.

4.1 Idiomatic use of virtual and acoustic instruments

In industry settings, decisions made around the use of virtual or acoustic instruments are largely governed by budget, the composer's own technical skills, and the wishes of the director (Ellis-Geiger, 2007). It can often be the case that a composer will imagine a cue that will sound effective when played using a real acoustic orchestra, but which may not, due to current technical limitations in virtual instrument

[2] The author is a professional composer, and has reviewed various composer interviews on the topic.

technology, sound acceptable using currently available virtual instruments. Conversely, it is possible to produce music cues using virtual acoustic instruments that are beyond the ability of human performers alone.

When crafting a string arrangement for real players, it is important to have a clear understanding of the kind of ensemble for which one is writing. A string ensemble can be anything from a string quartet to a full-sized string orchestra of 40 or more players and each instrument, and every player within the orchestra, has strengths and limitations. A composer must work within these parameters from the outset and the idiosyncrasies and particularities of each ensemble must be accounted for in the writing. In a similar sense, virtual instruments come with different strengths and weaknesses across their solo instrument, small ensemble, and large section libraries. The strengths and weaknesses arise not so much from the capability of the players, as from the contents and arrangement of sampled materials in each library and the way in which each library patch has been made responsive to human input.

It is important to understand the limitations of all instrumental forces, as these influence the composition of the cue or the arrangement. This applies to virtual instruments as much as it does to acoustic instruments. In addition, to create believable performances with virtual instruments, it is necessary to not just understand the intricacy of each instrument, but also intricacies of how an ensemble of musical performers will behave in a live context (e.g. individual variations in timing articulation intonation and so on). These subtleties need to be accounted for in the composition and production process.

4.2 Virtual instruments – strengths and weaknesses

If the objective is to construct a realistic simulation of acoustic instrumentation from virtual instruments then, given the above, the film score must be written specifically for the virtual acoustic instruments that will be used. Musical materials will be composed and developed that arise from a subset of instrumental gestures where the instruments sound most realistic. Instruments will exhibit weaknesses in certain areas. These weaknesses had to be identified and mitigated to ensure the most believable performances were recorded.

For example, virtual string instruments typically exhibit some consistent weaknesses. Pizzicato sequences, where the same note is plucked in quick sequence, result in an identical or very similar articulation presented in rapid succession, leading to the impression that a sample is being used and repeated for each note. Some software sample players (e.g. Native Instruments KONTAKT) try to mitigate this problem by randomising different samples of the same pitch for each articulated note, but the limited variation in articulation is still often noticeable. These weaknesses in a sample library become most acute when attempting to perform a solo instrument line that is exposed in the musical texture with little or no other musical material present.

In human performances of acoustic instruments, there is a large amount of nuance and complexity in most aspects of the production of notes and phrases. Few, if any, articulations of the same note sound the same. This complexity is present in all human performances of acoustic instruments. However, it is this complexity that is most difficult to emulate or reproduce when using a virtual instrument. Beyond micro-detail there is a range of different articulations associated with phrasing patterns. In string instruments this might include bow direction, bow position, legato, vibrato, glissando and so on. When an acoustic instrument is heard in isolation these details are clearly audible. Different compositions or music cues emphasise or foreground these elements to different degrees. Some forms of music depend on a complex presentation of these articulations as a core feature of the genre. In other genres or compositions, the musical materials are very ‘section based’ and such individual details are less noticeable due to the effect of aggregating many players. More ambient styles often focus on a reduced range of articulations and foreground pitch duration and density in preference to complex phrasing.

4.3 Conceptualization of music cues

Computer technology has added a new conceptual dimension to the process of film scoring, which has consequently transformed the process into an immensely integrative and technical art form. The film composer now conceptualizes their work through an intellectual process that integrates the physical workspace (the studio) and the virtual technologies available through computer applications, simulations and virtual instruments (Love, 2013, p. 13). When the composer reaches the ‘conceptualizing’ step in the scoring process, they must make several crucial musical decisions. It is that during this process that the composer establishes the foundational element of any good score: the score concept (Skelton, N.D.). As argued by Karolin and Wright, “the score concept is the primary idea that functions as a foundation upon which the score is built” (Karlin & Wright, 1990, p. 81). Its main purpose, regardless of the medium and the compositional methods used, is to connect the ideas, the environment and characters with music. At this point, the composer will gain a sense of the tone of the movie scene and will begin to picture the types of instruments, arrangements and music styles that can be provided by suitable textures.

4.4 Conceptualization of cues for this study

In the context of this study, the conceptualization of the score was to a certain extent determined by the research question, in that virtual instruments were used to mimic acoustic instruments. The emphasis was on producing music for virtual instruments, and in particular creating believable music using those instruments. As such, the cues were written and arranged in such a way so they would be effective when played using both the virtual and real orchestras. This required the cues to exhibit all the characteristics of real performers and, importantly, meant restricting the use of virtual instruments to material that is playable by human performers using acoustic instruments.

Financial constraints were also taken into account when deciding on how to compose the cues. With a very limited budget, it was known from the outset that hiring a 40-piece symphony orchestra was not realistic. Instead, the decision was made to use a five-piece acoustic ensemble, which consisted of an acoustic piano, two violins, viola and cello. However these five instruments provide vast range of feasible performance possibilities, both virtually and acoustically, and as such were seen as sufficient to investigate the research question.

In order to test the variables outlined in the methodology chapter, creative decisions were made regarding the selection of video examples, as well as the appropriate musical genres, instrumentations and musical arrangements for the selected video examples. The examples needed to allow for various aspects of the research questions to be explored, such as the extent to which visuals, sound fx and other tracklay elements might influence musical perception; the capacity for virtual instruments to convey emotion in a cue; and the extent to which genre or musical style might impact believability, since different genres involve different performance techniques. The two cues needed to provide enough scope to explore these dimensions.

4.5 Cue details

The research design called for the production of different versions of the two screen music cues so that the effect of different instrumentation and production process could be tested. Below are details of each of these cue versions and how they were organised and produced. The intention is also to outline the various technical and creative considerations that impacted both the cue content and the goal of achieving maximum believability in the use of virtual instruments.

4.5.1 Cue 1 Superman

The first video example from Superman Returns (2006) was selected to test the extent to which the presence of visual elements and other sounds (FX) affected the perception of believability of the listeners. It was also used to determine if a cue realized with virtual instruments for this particular movie scene is able to convey a strong emotional feeling. For this scene, the audience was to focus on the visuals, with the music acting in a supporting capacity. The Superman scene was an optimal choice, as it provided a dramatic and captivating visual scene upon which to focus. The action itself centres on Superman flying through a darkening city, high above the skyscrapers, before shooting off between the clouds and stars and into outer space. There, he hovers above earth, suspended without gravity while the camera zooms in on his face. The fast motion, combined with the dramatic scenery and additional sound effects, provide ample and sufficiently strong sensory material beyond the music track to serve as a platform to investigate the impact of surrounding materials on the perception of the music track in context.

4.5.2 Cue 2 Waltz

In contrast to the first scene, where music was acting in a supporting role, the goal for the second scene was to make the music the dominant and necessary feature. This was achieved by using the waltz scene from Shall We Dance (2004), whereby the music itself was a pivotal component of the setting. Unlike the Superman scene, the music in this scene can be best characterized as ‘diegetic’, which according to Chion is “a sound that imposes on the sequence a sense of real time, like normal everyday experience, and above all, a sense of time that is linear and sequential” (Chion, 1994, p. 18). The music is foregrounded in this video; it is one of the main sensory elements, not just in terms of being foregrounded in the mix, but also in terms of the action. In this video, a lack of music would have caused the audience to question its absence, so central was it to the understanding and context of the particular event in the film.

Additionally, the second scene provided greater dynamics, a faster pace and a much more lively musical and video context. In order to ensure the music was the dominant feature, a significantly different musical style was adopted in comparison to the first cue. The style was chosen to accompany the supporting images; it had to feel natural and in context. This different style also provided an opportunity to test the extent to which genre and instrumentation factors influence believability.

4.5.3 Cue 1 Superman – Virtual Instruments: Performed individually

The intention first and foremost was to produce an original piece of music that was well suited to convey the mood of the action sequence. The sequence shows Superman flying over darkening, brooding skies, over spectacular cityscapes and up float eerily in the stratosphere where he surveys a darkening planet. The mood of the sequence is sombre, introspective with an element of dream-like fantasy. The character displays little outward emotion, suggesting that he is reflecting deeply.

These qualities were underlined by the use of a cue in minor key with the main theme played by cello, an instrument with highly expressive, emotive potential. A series of slow dynamic-quality legato string samples were used along with a simple transition melody. The musical texture is open enough to allow the dramatic flying sound effects to be heard. There is little other sound other than the dramatic sound of his figure flying across the frame towards an eventual point of stillness. The music carries strong emotional purpose.

As explained in Chapter 4.2 common challenge with sampled string instruments is to effectively and accurately imitate the different articulations associated with the instrument. Real string instruments can articulate notes in a wide variety of ways – such as pizzicato, bowed - legato, spiccato, staccato, vibrato and so on. These all need to be effectively represented and made available in a virtual instrument context. This poses challenges from an instrument design perspective and so most libraries have weaknesses in some of these articulation areas.

In addition, due to the way in which samples are key-mapped[3] across different pitch ranges then certain notes can sound less realistic and pitch-shifted if they are too far away from the root sample note. As samples are pitch shifted upwards, the formant frequencies[4] of the sound is also shifted (unlike an acoustic instrument where the formant frequencies remain the same). When the sample is transposed upwards, this produces the well-known 'Mickey Mouse effect'.

Any given sample has a certain pitch range that sounds more realistic than the rest, therefore, finding the right register to play each sample was a key factor in achieving a realistic sound. This limitation is largely present due to the fact that most sample libraries (depending on size limitations) only have reference samples taken at intervals of 4ths, 5ths or an octave. The larger the note range a particular sample needs to cover, the more likely a sample will be played that is remote from its root pitch, thus the more likely it is to sound 'sampled'. Larger libraries tend to have more samples and so suffer less from this issue. Regardless, depending on the sample library and patch, some pitch ranges can sound more realistic than others. This is an important consideration for the composer when realism or believability are important.

Consequently, this cue has been shaped to a degree by the above limitations around articulation and pitch.

To effectively and accurately imitate from a MIDI controller legato playing styles, where each instrument was performed individually, it was necessary to assign (and gain control over) note attack and release times via a modulation wheel. In this

[3] Key mapping is a process of mapping different samples or instruments to different keys, octaves or regions on a MIDI keyboard so when each key is pressed a corresponding sample or instrument is played.
[4] Formants are the distinguishing or meaningful frequency components (or key resonant properties) that characterize the sound of an instrument. These can change in frequency depending on the size and shape of the instrument. Formant regions are not directly related to the pitch of the fundamental frequency and can remain constant as the fundamental changes (Sundberg, 1977).

instance, programming the modulation wheel to range from 0-sec attack/1-sec release at its minimum value to 0.5-sec attack/3-sec release at its maximum, the operation of the modulation wheel could alter the note envelope as the instrument was performed, naturally flowing between different note articulations - from detaché to legato. Incorporating this technique and adjusting the envelope of the sample made the performance expressive and more lifelike.

Another important effect produced by the string instruments in this cue was vibrato, a musical effect made up of a regular pulsating change of pitch. Choosing the right sample that was pre-recorded with vibrato was one option; the second option was to assign a 5-7Hz low frequency oscillator (LFO)[5] to a string patch in order to achieve a similar effect. Assigning different vibrato rates to each string instrument in the ensemble, as well as deeper modulation on the violins and less on the viola and the cello, made the string sound more realistic. In the arrangement, both violins play the same part throughout the song. Two different violin string patches were used and layered together, in doing so, the different patch textures created the impression of two different violin instruments (rather than the same instrument played twice) and so added to the realism of the cue.

As the only polyphonic instrument in this arrangement - the piano - provided the underlying harmonic content, and diversified the arrangement from a sonic perspective, Today's piano sample libraries bring unprecedented levels of realism. The piano sample library used in this cue was created in such way to capture every tonal characteristic of the real acoustic piano. Every key is recorded at multiple

[5] Low-frequency oscillation (LFO) is an electronic signal which is usually below 20 Hz and creates a rhythmic pulse or sweep. This pulse or sweep is often used to modulate synthesizers, delay lines and other audio equipment in order to create effects used in the production of electronic music. Audio effects such as vibrato, tremolo and phasing are examples. The abbreviation is also very often used to refer to low-frequency oscillators themselves.

dynamics levels, notes transition smoothly across the keyboard and through the dynamic range, with a full, rich tone throughout. Intelligent techniques were employed to avoid the evident 'sample switching' of other piano sounds to avoid unnatural behaviour when one piano sample is being changed to another. It offers pedalling and damper resonance and 256 note polyphony, harmonic resonance modelling for the most realistic sympathetic string resonance possible, audience, player, and close microphone recording positions and adjustable pedal noises. All these innovative implementations together made an extraordinary playing experience with an expressive range extending from the exuberant fortissimo to the most detailed and nuanced pianissimo.

In this version of the cue, all five instruments were performed individually in one take to a metronome using a MIDI keyboard for note input. As the dry samples do not indicate the presence of an acoustic space, a digital reverb was used. Quantizing was avoided. All original performance timings were retained in order to make the performance as realistic and similar to a real acoustic performance as possible.

4.5.4 Cue 1 Superman – Virtual instruments: Sequenced

The *second version* of the ambient cue was realized by manually entering and quantizing the notes in the arrangement out of real-time using a screen-based computer interface. In this version, there was no real-time human performances involved in the production of the cue whatsoever. While the virtual instruments playing the cue were the same, the performance factors contrasted greatly.

After the notes for each part in the arrangement were manually entered, each track was revised by adding automation; this was necessary in order to make the computerized performances sound more expressive. During a real ensemble performance, a musician cannot possibly maintain an exact volume. Using the expression envelope, the 'volume' was varied gently to introduce dynamic variation. In parts where a gentle introduction of a note or fade away was needed, bigger dips in the expression envelope were used; it was a subtle effect that added to realism. This is important because library string patches have dynamic layers (sample sets of made from instruments played at different dynamic levels), which fade between each other every time the modulation wheel is raised or lowered. Since string instruments have a vast expressive range, continuous controller (CC) data had to be used in order to portray this range.

The envelope was automated inside the sequencer (Cubase) as well as the vibrato effect and velocities. When performing live, a string player uses changes in bow direction as an integral part of the process of phrasing musical lines. It is an expressive musical parameter. Legato passages are played by incorporating a number of note articulations into one bow without any changes in bow direction. While some sample libraries provide legato articulations that are designed to remove the attack of the notes following the first, during this process the same effect was achieved by overlapping note events in a sequencer. The postproduction process for both versions of the cue involved adding sound effects such as city ambience, wind and swoosh sounds, panning, overall balancing of instruments and other sounds, and adding spatial effects.

4.5.5 Cue 2 Waltz – Virtual instruments: Performed individually

In contrast to the first Superman cue, the musical arrangement in this cue was more melodic and rhythmic; not only was the intention to make the music the main focus of the scene, it also had to work in conjunction with the supporting images. Due to the rhythmic timing of the dancing in the video, the tempo of the cue had to be adjusted to match the timing of the on screen action. The music ideally would sit in the foreground, just above the dialogue, atmospheres and foley, and would need to be mixed as if it where taking place in the actual ballroom.

Contrary to the production methods employed in the realization of the Superman cue, the production method of this cue didn't involve real time track automation, modulation of different aspects of the audio signals produced, quantization nor post editing of samples. Additionally, instrumental arrangement for this cue consisted of four instruments only: two violins, viola and cello. Acoustic piano was left out from this cue because it was not considered to be a realistic inclusion in this type of instrumental arrangement.

As with the Superman cue this version of the waltz cue was realized with all four instruments performed individually in one take using a MIDI keyboard for note input. Whilst a click track was used, quantizing was avoided. All original performance timings were retained in order to make the performance as realistic and similar to a real acoustic performance as possible. Reverberation was used during the performance to give a sense of performing in a live acoustic space.

4.5.6 Cue 2 Waltz – Virtual instruments: Sequenced

The second version of the waltz cue was realized by manually entering and quantizing the notes in the arrangement out of real-time using a screen-based computer interface. In this version, there was no real-time human performance involved in the production of the cue whatsoever. The virtual instruments playing both versions of the cue were the same. Just as with the first version of the waltz cue, the production method of this version of the cue didn't involve real time track automation, modulation of different aspects of the audio signals produced, nor post editing of samples. The postproduction process for both versions of the cue involved adding sound effects such as applause, footsteps, overall balancing of instruments and other sounds; and the addition of reverberation to give an impression that the music was being performed live by musicians in the scene.

4.5.7 Acoustic cues

The craft, and the challenge, of recording acoustic instruments is to capture the best possible interpretation of the musical performance without affecting the tonal quality of the instruments (Robjohns, 1999). During collaborative recording sessions such as this, it is crucial to achieve a positive and relaxed atmosphere in order to inspire and motivate the musicians to contribute artistically to the project. A fully notated score was prepared for all of the recording sessions and given to performers prior to the recording sessions. Virtual versions of the cues were also given to the musicians a week before the recording was scheduled so they could familiarize themselves with the material. Like most instruments, the sound produced by strings and acoustic piano preferably needs space to become properly balanced and coherent, therefore, appropriate recording techniques were employed in order to achieve successful recordings of both cues.

4.5.8 Cue 1 Superman – Acoustic ensemble: Performed together

To capture the natural tonal qualities of each instrument during the recording process the players were positioned in a relatively large environment that had fairly reflective surfaces with wood on both the floor and walls. The ensemble was recorded using a combination of close and ambient microphone techniques. Each string instrument was recorded using close microphone technique where microphones were positioned 0.5 m above the instruments, pointing down towards the respective instruments. An upright piano performed in this cue was recorded by positioning a stereo pair of microphones at the back of the instrument, 15 cm apart, at a distance of 1 m. The seating arrangement of a string quartet, looking at the stage from an audience perspective was in the order: first violin, cello, viola, and second violin. The arrangement was roughly an arc of a circle where the first violinist was almost facing the second violinist and the piano was positioned on the right side next to the second violin, 2 m away. Each performer was provided with a pair of headphones in order to allow for better communication between musicians, and between musicians and the composer/producer/engineer. To achieve a sense of live performance, the cues were performed and recorded in one take without overdubbing any of the parts. The sound of the instruments recorded was not altered in any way. Neither equalization nor compression was used in the postproduction process. The postproduction process involved adding only reverberation to overall mix of instruments.

4.5.9 Cue 1 Superman – Acoustic ensemble: Recorded multi-track

Performers were recorded in the same environment using the same recording technique. Each string instrument was recorded using a close microphone positioned above the instrument, (0.5 m) pointing down towards the instrument. An upright piano performed in this cue was recorded by positioning a stereo pair of microphones at the back of the instrument, 15 cm apart, at a distance of 1 m. Because the piano appears as the first instrument in the Superman cue, the acoustic piano was recorded first, followed by the cello, both violins and lastly the viola. The postproduction process involved overall balancing of instruments and adding reverberation.

4.5.10 Cue 2 Waltz – Acoustic ensemble: Performed together

As with the Superman cues, the ensemble was recorded in the same space using a combination of close and ambient microphone techniques. Each string instrument was recorded using a close microphone, positioned 0.5 m above each instrument, pointing down. In addition to the close microphones, a stereo XY[6] microphone pair was used in the middle of the recording space approximately 3 m away from the ensemble to capture the sound of the room. This recording technique was used in order to give the audience the impression that the recording took place in the actual ballroom where the scene takes place. A click track was used during the recording process since the music had to be synchronized with the images and the tempo of the on screen action (dance). During postproduction, the recorded sound of the instruments was not altered in any way. Neither equalization nor compression was used. The postproduction process involved balancing the sound of music with other sounds in picture and adding reverberation.

4.5.11 Cue 2 Waltz – Acoustic ensemble: Recorded multi-track

A similar recording approach was used for recording the multi-tracked version of the waltz cue. Each of four string instruments was recorded using close microphone positioned above the instrument (0.5 m) and pointing down towards the instrument. Additionally, each instrument was recorded using a stereo XY microphone pair positioned in the middle of the recording space at a distance of 1.5 m to capture a more detailed recording. During the process of postproduction the recorded sound of the instruments was not altered in any way. Neither equalization nor compression was used. The postproduction process involved balancing the sound of music with other sounds in picture and adding reverberation.

[6] Stereo recording technique where two coincident directional microphones are used at and angle of 90° .

4.5.12 Conclusion

This chapter explained how the composer-researcher produced the cues in order to achieve what was thought to be the highest levels of believability, prior to testing the results on listeners. It analysed the strengths and identified the potential weaknesses of virtual instruments to ensure the most believable performances were produced. This chapter examined the different composition and production approaches that, together with the listener responses outlined in the next chapter, will provide insights into the extent to which these might influence believability.

Chapter 5: Discussion

The research has been designed to generate and test a broad range of possible variables in the production and reception stages of a music cue. These variables have been created to test the extent to which they might impact the believability of a given cue. Groups of questions were oriented towards exploring the following variables in the context of virtual and acoustic instruments:

- Performance and production factors (ensemble, solo multi-tracked performance, individual MIDI keyboard performance, mouse based note entry)
- Genre/style
- Presence or absence of visual elements and sound effects
- Musical background of respondent
- Emotional feeling

Lastly, the listening test was also designed to determine if the participant had a preferred musical cue and if so, whether a pattern emerged for a preference towards real or virtual instruments.

5.1 Performance and production factors

The test data showed that performance and production factors were not significant[7] in the listeners' perception of Superman and Waltz acoustic cues when cues were played in the presence of visuals and other sound effects. The same proportion of listeners identified examples as 'virtual' or 'acoustic' regardless of whether the cues were constructed via real time performances or multi-tracking (Table 5).

---

[7] The term "significant" in survey context refers to the fact that in a sample of 36 participants more people picked a particular example, as real or synthetic, the term "significant" does not refer to the technical definition of 'statistical significance'.

		Played examples			
		Superman		Waltz	
		Acoustic ensemble: Performing together	Acoustic ensemble: Recorded multi-track	Acoustic ensemble: Performing together	Acoustic ensemble: Recorded multi-track
Video format	Acoustic	88.9%	88.9%	88.6%	91.7%
	Virtual	11.1%	8.3%	8.6%	5.5%
	Mixture	0.0%	2.8%	2.8%	2.8%
Total		100.0%	100.0%	100.0%	100.0%

Table 5. Summary percentage data of listeners' perceptions Superman and Waltz acoustic cue versions (music played with video).

The same pattern emerged when acoustic cues were played in isolation from other sounds and visual elements (Table 6).

		Played examples			
		Superman		Waltz	
		Acoustic ensemble: Performing together	Acoustic ensemble: Recorded multi-track	Acoustic ensemble: Performing together	Acoustic ensemble: Recorded multi-track
Audio format	Acoustic	91.6%	94.4%	88.6%	91.7%
	Virtual	5.6%	5.6%	2.8%	5.5%
	Mixture	2.8%	0.0%	8.6%	2.8%
Total		100.0%	100.0%	100.0%	100.0%

Table 6. Summary percentage data of listeners' perceptions Superman and Waltz acoustic cue versions (music played without video).

The test data showed that performance and production factors were not significant in the listeners' perception of Superman and Waltz virtual cues when cues were played in the presence of visuals and other sound effects. The same proportion of listeners identified examples as 'virtual' or 'acoustic' regardless of whether the cues were constructed via real time performances or sequencing (Table 7).

		Played examples			
		Superman		Waltz	
		Virtual instrument: Performed individually	Virtual instrument: Sequenced	Virtual instrument: Performed individually	Virtual instrument: Sequenced
Video format	Acoustic	86.1%	83.3%	88.6%	88.9%
	Virtual	11.1%	11.1%	8.6%	8.3%
	Mixture	2.8%	5.6%	2.9%	2.8%
Total		100.0%	100.0%	100.0%	100.0%

Table 7. Summary percentage data of listeners' perceptions of Superman and Waltz virtual acoustic cue versions (music played with video)

The same pattern emerged when virtual acoustic cues were played in isolation from other sounds and visual elements (Table 8).

		Played examples			
		Superman		Waltz	
		Virtual instrument: Performed individually	Virtual instrument: Sequenced	Virtual instrument: Performed individually	Virtual instrument: Sequenced
Audio format	Acoustic	44.4%	41.6%	41.6%	41.6%
	Virtual	44.4%	44.4%	50.0%	47.2%
	Mixture	11.1%	13.9%	8.3%	11.1%
Total		100.0%	100.0%	100.0%	100.0%

Table 8. Summary percentage data of listeners' perceptions of Superman and Waltz virtual acoustic cue versions (music played without video).

5.2 Presence or absence of visual elements and sound effects

The test data showed that the presence or absence of visual elements and sound effects did not significantly affect the listeners' perception of Superman and Waltz acoustic cue versions. The same proportion of listeners identified examples as 'virtual' or 'acoustic' regardless of whether the cues were played in the context of other sounds and pictures, or in isolation from other sounds and visual elements (Table 9).

		Played examples			
		Superman		Waltz	
		Acoustic ensemble: Performing together	Acoustic ensemble: Recorded multi-track	Acoustic ensemble: Performing together	Acoustic ensemble: Recorded multi-track
Audio format	Acoustic	91.6%	94.4%	88.6%	91.7%
	Virtual	5.6%	5.6%	2.8%	5.5%
	Mixture	2.8%	0.0%	8.6%	2.8%
Video format	Acoustic	88.9%	88.9%	88.6%	91.7%
	Virtual	11.1%	8.3%	8.6%	5.5%
	Mixture	0.0%	2.8%	2.8%	2.8%
Total		100.0%	100.0%	100.0%	100.0%

Table 9. Summary percentage data of listeners' perceptions Superman and Waltz acoustic cue versions (four of which are in Audio format, and four in Video format).

The test data showed that the presence or absence of visual elements and sound effects was a significant factor in the listeners' perception of Superman and Waltz virtual acoustic cues (Table 10).

		Played examples			
		Superman		Waltz	
		Virtual instrument: Performed individually	Virtual instrument: Sequenced	Virtual instrument: Performed individually	Virtual instrument: Sequenced
Audio format	Acoustic	44.4%	41.6%	41.6%	41.6%
	Virtual	44.4%	44.4%	50.0%	47.2%
	Mixture	11.1%	13.9%	8.3%	11.1%
Video format	Acoustic	86.1%	83.3%	88.6%	88.9%
	Virtual	11.1%	11.1%	8.6%	8.3%
	Mixture	2.8%	5.6%	2.9%	2.8%
Total		100.0%	100.0%	100.0%	100.0%

Table 10. Summary percentage data of listeners' perceptions of Superman and Waltz virtual acoustic cue versions (four of which are in Audio format, and four in Video format).

5.3 Genre/style

The test data showed that genre/style factors were not significant in the listeners' perception of Superman and Waltz acoustic cues when cues were played in the presence of visual elements and other sound effects. The same proportion of listeners identified examples as 'virtual' or 'acoustic' regardless of whether the cues were Superman or Waltz (Table 11).

		Played examples			
		Superman		Waltz	
		Acoustic ensemble: Performing together	Acoustic ensemble: Recorded multi-track	Acoustic ensemble: Performing together	Acoustic ensemble: Recorded multi-track
Video format	Acoustic	88.9%	88.9%	88.6%	91.7%
	Virtual	11.1%	8.3%	8.6%	5.5%
	Mixture	0.0%	2.8%	2.8%	2.8%
Total		100.0%	100.0%	100.0%	100.0%

Table 11. Summary percentage data of listeners' perceptions Superman and Waltz acoustic cue versions (music played with video).

The same pattern emerged when acoustic cues were played in isolation from other sounds and visual elements (Table 12).

		Played examples			
		Superman		Waltz	
		Acoustic ensemble: Performing together	Acoustic ensemble: Recorded multi-track	Acoustic ensemble: Performing together	Acoustic ensemble: Recorded multi-track
Audio format	Acoustic	91.6%	94.4%	88.6%	91.7%
	Virtual	5.6%	5.6%	2.8%	5.5%
	Mixture	2.8%	0.0%	8.6%	2.8%
Total		100.0%	100.0%	100.0%	100.0%

Table 12. Summary percentage data of listeners' perceptions Superman and Waltz acoustic cue versions (music played without video).

The test data showed that genre/style factors were not significant in the listeners' perception of Superman and Waltz virtual acoustic cues when cues were played in the presence of visual elements and other sound effects. The same proportion of listeners identified examples as 'virtual' or 'acoustic' regardless of whether the cues were Superman or Waltz (Table 13).

		Played examples			
		Superman		Waltz	
		Virtual instrument: Performed individually	Virtual instrument: Sequenced	Virtual instrument: Performed individually	Virtual instrument: Sequenced
Video format	Acoustic	86.1%	83.3%	88.6%	88.9%
	Virtual	11.1%	11.1%	8.6%	8.3%
	Mixture	2.8%	5.6%	2.9%	2.8%
Total		100.0%	100.0%	100.0%	100.0%

Table 13. Summary percentage data of listeners' perceptions of Superman and Waltz virtual acoustic cue versions (music played with video)

The same pattern emerged when virtual acoustic cues were played in the absence of video elements and other sound affects (Table 14).

		Played examples			
		Superman		Waltz	
		Virtual instrument: Performed individually	Virtual instrument: Sequenced	Virtual instrument: Performed individually	Virtual instrument: Sequenced
Audio format	Acoustic	44.4%	41.6%	41.6%	41.6%
	Virtual	44.4%	44.4%	50.0%	47.2%
	Mixture	11.1%	13.9%	8.3%	11.1%
Total		100.0%	100.0%	100.0%	100.0%

Table 14. Summary percentage data of listeners' perceptions of Superman and Waltz virtual acoustic cue versions (music played without video).

5.4 Musical background of respondent

The test data showed that the musical background of respondents was not a significant factor in the listeners' perception of Superman and Waltz acoustic cues when cues were played in the presence of visuals and other sound effects. The same proportion of listeners identified examples as 'virtual' or 'acoustic' regardless of their musical background. The same pattern emerged when Superman and Waltz acoustic cues were played in isolation from other sounds and visual elements (Table 15).

Ability to play musical instrument →		Played Example			
		Acoustic ensemble: Recorded multi-track		Acoustic ensemble: Performing together	
		Do not play	Play	Do not play	Play
Superman audio format	Acoustic	84.20%	94.10%	89.50%	88.20%
	Virtual	10.50%	5.90%	10.50%	11.80%
	Mixture	5.30%	0.00%	0.00%	0.00%
Superman video format	Acoustic	89.50%	100.00%	89.50%	94.10%
	Virtual	10.50%	0.00%	5.30%	5.90%
	Mixture	0.00%	0.00%	5.30%	0.00%
Waltz audio format	Acoustic	94.70%	88.20%	84.20%	94.10%
	Virtual	5.30%	5.90%	10.50%	5.90%
	Mixture	0.00%	5.90%	5.30%	0.00%
Waltz video format	Acoustic	89.50%	88.20%	89.50%	94.10%
	Virtual	10.50%	5.90%	5.30%	5.90%
	Mixture	0.00%	5.90%	5.30%	0.00%

Table 15. Summary data of listeners' perceptions of Superman and Waltz acoustic cues when cues were played in the presence or in isolation from other sounds and visual elements. The results are resolved according to whether the listener plays or does not play a musical instrument.

The test data showed that the musical background of respondents was not a significant factor in the listeners' perception of Superman and Waltz virtual cues when cues were played in the presence of visuals and other sound effects. The same proportion of listeners identified examples as 'virtual' or 'acoustic' regardless of their musical background (Table 16).

Ability to play musical instrument →		Played Example			
		Virtual instrument: Sequenced		Virtual instruments: Performed individually	
		Do not play	Play	Do not play	Play
Superman video format	Acoustic	84.20%	82.30%	84.20%	88.20%
	Virtual	10.50%	11.70%	15.70%	5.90%
	Mixture	5.30%	5.90%	0.00%	5.90%
Waltz video format	Acoustic	84.20%	94.10%	84.20%	82.40%
	Virtual	10.50%	5.90%	15.80%	5.80%
	Mixture	5.30%	0.00%	0.00%	11.80%

Table 16. Summary data of listeners' perceptions of Superman and Waltz virtual cues when cues were played in the presence of visuals and other sound effects. The results are resolved according to whether the listener plays or does not play a musical instrument.

The test data showed that the musical background of the respondent was a significant factor in the perception of Superman and Waltz virtual cues when cues were played in isolation from other sounds and visual elements (Table 17).

Ability to play musical instrument →		Played Example			
		Virtual instrument: Sequenced		Virtual instruments: Performed individually	
		Do not play	Play	Do not play	Play
Superman audio format	Acoustic	63.20%	20.00%	63.10%	23.50%
	Virtual	31.60%	66.70%	26.30%	64.70%
	Mixture	5.30%	13.30%	10.50%	11.80%
Waltz audio format	Acoustic	57.80%	23.50%	52.60%	29.40%
	Virtual	21.10%	76.50%	36.80%	64.70%
	Mixture	21.10%	0.00%	10.50%	5.90%

Table 17. Summary data of listeners' perceptions of Superman and Waltz virtual cues when cues were played in isolation from other sounds and visual elements. The results are resolved according to whether the listener plays or does not play a musical instrument.

5.5 Emotional feeling

The test data showed that both Superman and waltz cues were able to convey strong emotional feeling regardless of being played in the presence or absence of visual elements and sound effects and regardless of the instrumentation or production process employed (Tab.18, Tab.19, Tab.20, Tab.21).

		Played Versions			
		Virtual instrument: Performed individually	Acoustic ensemble: Recorded multi-track	Virtual instrument: Sequenced	Acoustic ensemble: Performing together
Strength	Weak	16.7%	2.8%	11.1%	2.8%
	Strong	83.3%	97.2%	88.9%	97.2%
Total		100.0%	100.0%	100.0%	100.0%

Table 18. Cue 1 Superman - Emotional strength (music played with video)

		Played Versions			
		Virtual instrument: Performed individually	Acoustic ensemble: Recorded multi-track	Virtual instrument: Sequenced	Acoustic ensemble: Performing together
Strength	Weak	0.0%	0.0%	0.0%	2.9%
	Strong	100.0%	100.0%	100.0%	97.1%
Total		100.0%	100.0%	100.0%	100.0%

Table 19. Cue 2 Waltz - Emotional strength (music played with video)

		Played Versions			
		Virtual instrument: Performed individually	Acoustic ensemble: Recorded multi-track	Virtual instrument: Sequenced	Acoustic ensemble: Performing together
Strength	Weak	0.0%	5.7%	2.9%	0.0%
	Strong	100.0%	94.3%	97.1%	100.0%
Total		100.0%	100.0%	100.0%	100.0%

Table 20. Cue 1 Superman – Emotional strength (music played without video)

		Played Versions			
		Virtual instrument: Performed individually	Acoustic ensemble: Recorded multi-track	Virtual instrument: Sequenced	Acoustic ensemble: Performing together
Strength	Weak	2.9%	0.0%	5.7%	2.9%
	Strong	97.1%	100.0%	94.3%	97.1%
Total		100.0%	100.0%	100.0%	100.0%

Table 21. Cue 2 Waltz - Emotional strength (music played without video)

5.6 Preferred example

The test data showed that the participants had an overall preference for examples made with an acoustic ensemble and displayed the strongest preference for acoustic cues realized through multi-tracking (Tab. 22, Tab. 23, Tab. 24, Tab. 25,).

Played examples	Count	Percent
Acoustic ensemble: Performing together	10	27.70%
Acoustic ensemble: Recorded multi-track	12	33.30%
Virtual instrument: Sequenced	6	16.70%
Virtual instrument: Performed individually	8	22.20%
Total	36	100%

Table 22 Summary data of listeners' preferred example from the four Superman cues (music played with video).

Played examples	Count	Percent
Acoustic ensemble: Performing together	11	30.60%
Acoustic ensemble: Recorded multi-track	12	33.30%
Virtual instrument: Sequenced	9	25.00%
Virtual instrument: Performed individually	4	11.10%
Total	36	100%

Table 23. Summary data of listeners' preferred example from the four Waltz cues (music played with video).

Played examples	Count	Percent
Acoustic ensemble: Performing together	9	25.00%
Acoustic ensemble: Recorded multi-track	15	45.50%
Virtual instrument: Sequenced	5	13.90%
Virtual instrument: Performed individually	7	19.40%
Total	36	100%

Table 24. Summary data of listeners' preferred example from the four Superman cues (music played without video).

Played examples	Count	Percent
Acoustic ensemble: Performing together	11	30.60%
Acoustic ensemble: Recorded multi-track	12	33.30%
Virtual instrument: Sequenced	7	19.40%
Virtual instrument: Performed individually	6	16.70%
Total	36	100%

Table 25. Summary data of listeners' preferred example from the four Superman cues (music played without video).

5.7 Overall observations

Reviewing the results, what is immediately obvious is that on no occasion are all 36 participants unanimous in their perception of a given piece of music. That is, in no scenario do all listeners perceive a cue as being either 'acoustic' or 'virtual acoustic'.

A thorough analysis of the data draws out some recurring patterns. First, no factor had a significant impact on the listeners' perception of the acoustic or virtual sources when cues were played in the presence of visual elements and sound effects.

Second, participants were better able to identify cues made with virtual instruments when played without visual elements and sound effects, i.e. as audio only. This suggests that without the presence of visual elements and sound effects the participants were able to better focus on the sounds they were hearing and consequently heard differences. Furthermore, participants with a musical background were more able to correctly identify virtual instruments when played without visual elements and other sound effects. This is the only instance where musical background appeared to impact believability. A review of the corresponding open-ended responses suggests that these participants have trained ears and were able to identify subtle nuances in the cue that participants without musical background failed to hear (Table 26).

“I don’t think that the music was made with real orchestra because the way instruments were performed seemed artificial.” “It sounds really nice, almost too perfect and clean to be recorded from real orchestra so I would say virtual.” “It sounds as if someone plays every instrument on a keyboard piano.” “The sound of the violin didn’t sound like it was violin.” “It didn’t sound real, the way the instruments were playing together especially cello in the beginning was very flat and boring.” “It did not feel real because there was no dynamic in the performance.”

Table 26 Selected responses from listeners with a musical background to Superman and Waltz cues made using virtual instruments with no visual elements present

Based on these observations, it can be concluded that the most significant factor was the impact of visual and other sound effects on the listeners’ perception of cues made with virtual instruments. This can be best illustrated by reviewing some of the open-ended responses of listeners in Table 27.

“Music and picture together made me think it was acoustic.” “With picture, focus is shifted elsewhere and this sounds like acoustic instrument and fits the bill.” “I am not sure, It felt like it was acoustic, I wasn’t paying attention to the music that much.”

Table 27 Selected listener open-ended responses to Superman and Waltz cues in the presence of visuals and other sound effects

Third, participants considered both virtual cues to be emotional even when they did not perceive them to be acoustic. For example, in the case of “Waltz Cue 2 virtual instruments performed individually”, only 41.6% of participants identified the track as acoustic (Table 10). Despite this, 97.1% of participants rated this example as having a strong emotional feeling (Table 21).

Notwithstanding the small scale of this study (36 participants) and the associated margins for error[8], the test data have still demonstrated that visual elements impact how we hear and perceive music. This is particularly so in the context of virtual acoustic instruments, irrespective of the musical genre, performance and production techniques, or the musical background of the audience. This finding corresponds with those of Wingstedt, Brändström, & Berg, who similarly found that visual elements of multimedia such as videos and computer games impact how we hear music (Wingstedt, Brändström, & Berg, 2010, p. 2).

Surprisingly, only a very small percentage of participants gave responses to the open ended questions in the questionnaire. This is possibly explained by the fact that participants either didn't have any comments, didn't feel their comments would add any value, or were unsure about what to say. The tendency to avoid these questions was the same between both participants from musical and non-musical backgrounds, indicating it had little to do with musical knowledge. The wording of the question perhaps also made it difficult for respondents to formulate a response. Another unexpected result was that none of the acoustic cues were perceived by all participants as being acoustic. The percentage of people who perceived the acoustic cues as being 'virtual' ranged from 2.8% in the case of "Cue 2 Waltz Acoustic Ensemble: performed together" (Table 11), to 11.1 % in the case of " Cue 1 Superman: Acoustic Ensemble Performed Together" (Table 5). A review of the

[8] The **margin of error** is a data indicating the amount of random sampling error in any survey result. The larger the margin of error, the less accurate the survey results are. It occurs whenever a population is incompletely sampled. *The margin of error is* calculated based on the formula: *The margin of error in a sample = 1 divided by the square root of the number of people in the sample.*

corresponding open-ended responses suggest this could be explained by a lack of certainty as to what participants thought they were listening to, or an inability to adequately describe their listening experience, as outlined in Table 28.

Table 28 Selected listener open-ended responses when asked to describe their listening experience

"I don't know, maybe it's virtual"
"Not sure"
"Sounds like"
"It sounded virtual, that's it"

Ultimately, what the results highlight, is that in most contexts the virtual acoustic instruments were considered believable to the listeners. In the context of this study, we should consider believability to be a pattern or trend rather than an absolute concept. As indicated above, the data provides no absolute results on this issue. What that pattern uncovered was that when virtual instruments were played in conjunction with visual elements and other sound effects, a significant proportion of the given sample, 86%, experienced virtual acoustic instruments as being acoustic. As such, it is fair to suggest that the virtual instruments were believable to a high proportion of the audience. Another trend that appeared in the results was that when virtual acoustic instruments were played in the absence of visual elements and other sound effects then 42.3% of the time participants experienced virtual acoustic instruments as being acoustic. This translates as roughly half of the time participants heard these cues as samples, meaning they cannot be considered believable, or at least, were less believable.

Overall, a review of the test data shows that the overarching trend of this study was that 86% of participants experienced virtual acoustic instruments as being acoustic. It is important to factor, however, the way in which the cues were conceptualized, composed and produced. This study deliberately includes the composition and production stage within the field of examination and as such illuminates the way in which cues were crafted in such way as to make them sound believable to the composer prior to the listening tests occurring. In effect, the composer had already made efforts to produce ‘realistic’ imitations of acoustic instruments from virtual cues by orientating the arrangement in each cue towards the strengths (and away from the weaknesses) of virtual instruments. This is an important contextual factor when one examines the data from the listening tests. The intention here is to include creative factors that are typical in a professional film scoring process.

As a consequence, when one cites the data from a study of this nature on might say that in the context where a composer has optimised a virtual instrument music cue for maximum believability, the study found that 86% of people perceived the virtual instruments as acoustic. This is *not* to imply that 86% of people will perceive virtual acoustic instruments as acoustic in all contexts. Rather, this will be true if composers optimise cues for maximum believability in the composition and production phase.

5.8 Limitations of the study

This study was limited in scope and scale by virtue of it being a master’s project with a limited project timeline and limited access to resources. As such, the findings from the project should be read in this context. The following section outlines the study limitations.

First, creating the cues with real musicians was necessary in order to establish a control version against which virtual cues could be compared when analysing listener responses. The project's limited budget meant, however, that it was not possible to hire a large orchestral ensemble. Composing music for larger orchestral ensembles would allow for the use of different instrumentation and more complex musical arrangements that may yield different results. The project is therefore limited in scope to a study of small ensemble contexts and the results cannot be generalised to larger ensembles.

Second, there was only one composer involved. Given that believability begins with the composer conceptualising cues in respect of the available virtual and acoustic resources (and their relative strengths and weaknesses) the composition phase represents a variable in the system that requires more extensive testing and examination. A larger study involving multiple composers would shed more detailed light on the impact of the composer's craft and approach on the believability of the end result. Indeed, such an increase in scale would also allow for the testing of a broad range of virtual instrument libraries as a further variable. Due to the limited scope of this study, only a select group of virtual instruments could be deployed.

Third, due to the time and resource constraints of the study, a limited number of participants could be surveyed (36 participants). A larger sample would provide more robust data, lower the margin of error in responses, and the survey data would become more reliable.

Fourth, the timeframe for the execution of the research study was restricted to eight months. A longer timeframe, such as that granted for a PhD thesis, would allow for the production of a larger quantity of cues for the listening test phase. A greater number of stylistically different versions would allow for believability to be tested across different musical genres, different video contexts, and different arrangement approaches.

Despite these limitations, the study yields a range of findings that provide a solid platform to design larger, more comprehensive research projects that build upon and explore more deeply the identified questions.

Chapter 6: Conclusion

This thesis has concerned itself with an examination of the factors that impact the believability of virtually orchestrated film scores. To best understand these factors, however, we had to understand first-hand through practice-led research the way in which the compositional process might influence this investigation and then explore the perception of typical cinema audience members through a structured listening test.

Chapters 1 and 2 outlined the research question, examples of previous research and discourses on believability, as well as historical attempts to analyse this form of music and technology. This provided a context in which to place the current research and the challenges associated with analysing and composing with virtual instruments overall.

Chapter 3 presented an overview of the research methodology and described the research design of the study in order to demonstrate that the research methods employed were both adequate and suitable to answering the research questions.

Chapter 4 provided insight and reflection of the creative practice from a composer's perspective. It explained how the composer-researcher produced the cues in order to achieve what was thought to be the highest levels of believability, prior to testing the results on listeners. Strengths and potential weaknesses of virtual instruments were identified to ensure the most believable performances were created and these were outlined and discussed. The chapter examined the different composition and production approaches that could be correlated to the listener responses outlined in

Chapter 5, in order to examine which, if any, of these factors might impact believability.

Finally, Chapter 5 provided the results of the research and a discussion of the most prominent and interesting results, the most significant factors impacting believability, and the overarching trends of the study. It drew conclusions and pointed out how the research results should be interpreted within the context of the study in order to understand their significance. The findings of this study indicated that:

- Performance and production factors were not significant in the listeners' perception of Superman and Waltz virtual or acoustic cues regardless of whether the cues were played in the presence or absence of visual elements and other sound effects.
- Genre/style factors were not significant in the listeners' perception of Superman and Waltz virtual or acoustic cues regardless of whether the cues were played in the presence or absence of visual elements and other sound effects.
- Presence or absence of visual elements did not significantly affect the listeners' perception of Superman and Waltz acoustic cues versions.
- Presence or absence of visual elements was a significant factor in the listeners' perception of Superman and Waltz virtual acoustic cues.
- Musical background of the respondent was not a significant factor in the listeners' perception of Superman and Waltz acoustic cues regardless of whether the cues were played in the presence or absence of visual elements and other sound effects.

- Musical background of the respondent was not a significant factor in the listeners' perception of Superman and Waltz virtual cues when cues were played in the presence of visual elements and other sound effects.
- Musical background of the respondent was a significant factor in the perception of Superman and Waltz virtual cues when cues were played in isolation from other sounds and visual elements.
- Both Superman and waltz cues were able to convey strong emotional feeling regardless of whether the cues were played in the presence or absence of visual elements and sound effects and regardless of the instrumentation or production process employed.
- The participants had an overall preference for examples made with an acoustic ensemble and displayed the strongest preference for acoustic cues realized through multi-tracking.

6.1 Significance of this research

The literature review for this project indicated that this area of practice is yet to receive systematic scholarly attention. There are no precedent studies and no specific sources in literature that directly address these research questions. The use of virtual instruments is pervasive in all professional screen music scoring contexts and yet little is known about the factors that impact believability, beyond the anecdotal observations of composers in the field. As such, this project constitutes an important initial contribution to the diverse community of educators, researchers and practitioners who are working closely with new technologies in the fields of music technology education, music composition, and the film industry. Researchers, musicians and composers will benefit from a better understanding of how computer technology and virtual acoustic instruments can be used to produce a realistic film scoring. The findings will be useful to inform a film composer's process where creative decisions are consistently being made around when to use real acoustic orchestras and when to use virtual acoustic instruments. This research will provide us with a systematic way of making these decisions and to understand more clearly the benefits and limitations of various approaches.

6.2 Opportunities for further research

As indicated above in Section 5.8, this study has been limited in scope due to the constraints of scale, time and resources imposed by a master’s project. Nevertheless, the study identified a number of areas that invite much closer investigation. These include a more detailed examination of the creative processes of multiple composers to understand the common principles (if any) that underpin their attempts to produce believable cues with virtual instruments. A larger study would permit this.

If more resources were available, it would be possible to broader the scope of the materials under test to include larger ensembles, orchestras and a much broader variety of instrumentation. It would also be possible to produce a higher number of test cues in a broader range of musical styles and with a broader ranger of visual materials. This would provide a more comprehensive set of test materials that would allow for more detailed understandings to be gained around the impact of instrumentation, genre/style and visual context on believability.

A larger study is needed to increase the scale of testing program so that it includes more listeners. Testing the materials on 100 or more listeners would start to refine and strengthen the data set to permit more rigorous and defensible conclusions.

Furthermore, a larger study might permit a systematic study of contrasting groups of listeners from musical and non-musical backgrounds. A small study, whilst showing some sign of trends, does not yield sufficient data to make robust comparisons between these two demographic groups. It is a reasonable hypothesis that musicians are more expert listeners who may identify virtual instruments more consistently and so a larger sample size would permit an initial examination of this factor.

What is clear is that this study opens up significant areas for further research by providing an initial pilot study that might inform much larger research efforts.

Appendix – Listening Test Data

The aggregated results of the listening test are presented in the form of tables and figures/charts below. There is one table of numerical results and one corresponding chart, representing the data for each multiple choice question. Responses to open ended questions have not been tabulated but are summarised, used and referenced where appropriate. In addition to a table of numerical results and corresponding chart, which represent the data for each multiple choice question, this data summary also includes one table of numerical results and one corresponding chart, which represent participant responses when music was played in the presence of visual elements and other sound effects compared to those played without. This chapter also includes summarized data comparing the responses of the participants with an ability to play musical instruments and those without, in order to test the impact of musical background on the perception of believability.

There were 36 participants involved in the experiment. The listening cohort were played a total of 16 different listening examples as explained more fully in Chapter 2 of this thesis. The tables and charts below represent aggregated views of the questionnaire response data and present them in an organised manner.

A.1 Listening test results – data summary

A.1.1 Cue 1 – Superman (music played with video)

		Played Versions			
		Virtual instrument: Performed individually	Acoustic ensemble: Recorded multi-track	Virtual instrument: Sequenced	Acoustic ensemble: Performing together
	Acoustic	86.1%	88.9%	83.3%	88.9%
Perception	Virtual	11.1%	8.3%	11.1%	11.1%
	Mixture	2.8%	2.8%	5.6%	0.0%
Total		100.0%	100.0%	100.0%	100.0%

Table 29. Cue 1 Superman (music played with video) - Listener perception identification summary

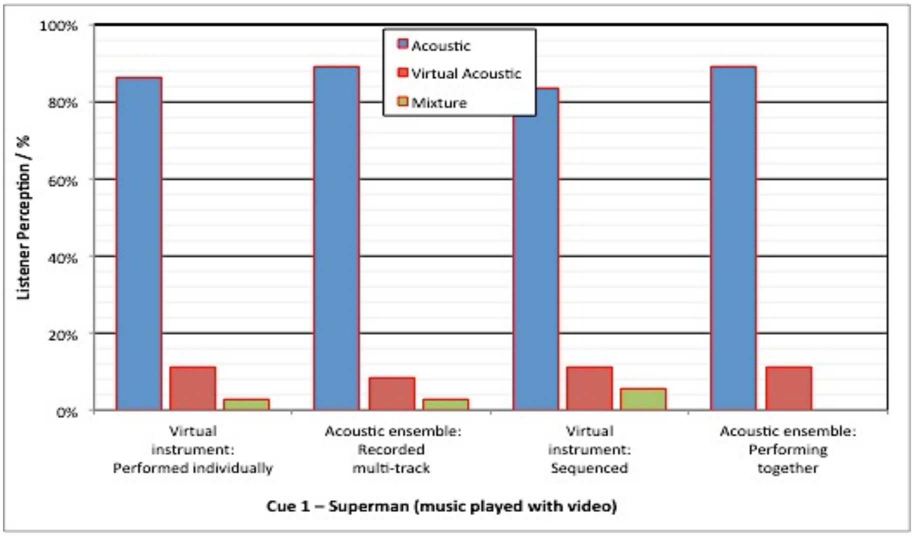

Figure 1. Plot of the summary percentage data form Table 29 for Cue 1 Superman (music played with video).

Table 29 shows the listener responses to the four different versions of Cue 1 Superman (**Question 1**), where the music cue is played with the video material and surrounding audio effects. The played versions are indicated and the listener data concerning source identification is presented underneath each version. For each example, listeners were asked to identify the source as ‘acoustic’, ‘virtual acoustic’ or ‘mixture’. **Figure 1** charts these responses graphically.

	Played Versions			
	Virtual instrument: Performed individually	Acoustic ensemble: Recorded multi-track	Virtual instrument: Sequenced	Acoustic ensemble: Performing together
Strength — Weak	16.7%	2.8%	11.1%	2.8%
Strength — Strong	83.3%	97.2%	88.9%	97.2%
Total	100.0%	100.0%	100.0%	100.0%

Table 30. Cue 1 Superman - Emotional strength (percentage representation).

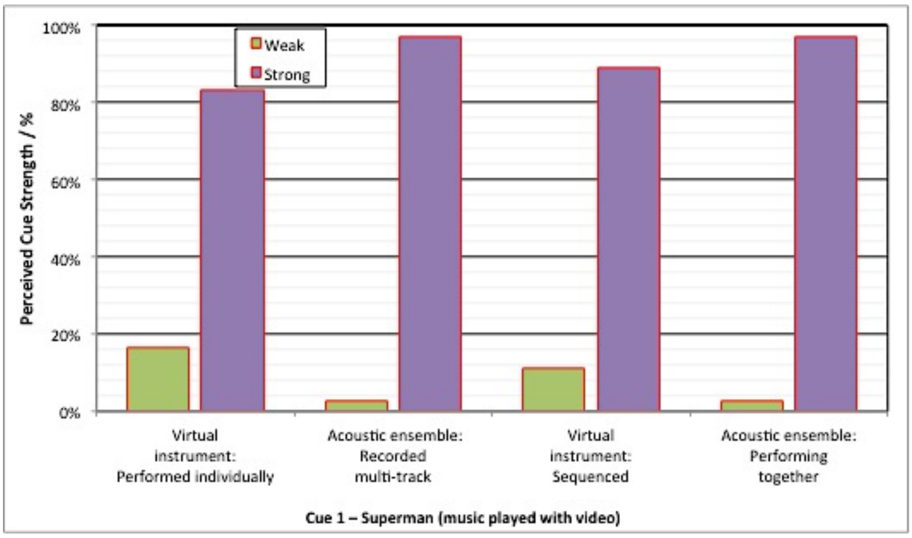

Figure 2. Cue 1 Superman - Emotional strength (graphic representation).

Table 30 shows the listener responses to the four different version of Cue 1 Superman **(Question 3),** where the music cue was played with the video material and surrounding audio effects. The played versions are indicated and the listener data concerning source identification is presented underneath each version. For each example, listeners were asked to identify the emotional strength of the source as 'acoustic', 'strong or 'weak'. **Figure 2** charts these responses graphically.

A.1.2 Cue 2 - Waltz (music played with video)

		Played Versions			
		Virtual instrument: Sequenced	Acoustic ensemble: Performing together	Virtual instrument: Performing individually	Acoustic ensemble: Recorded multi-track
Perception	Acoustic	88.9%	88.6%	86.1%	91.7%
	Virtual	8.3%	8.6%	8.3%	5.5%
	Mixture	2.8%	2.9%	5.6%	2.8%
Total		100.0%	100.0%	100.0%	100.0%

Table 31. Cue 2 Waltz (music played with video) - Listener perception identification summary.

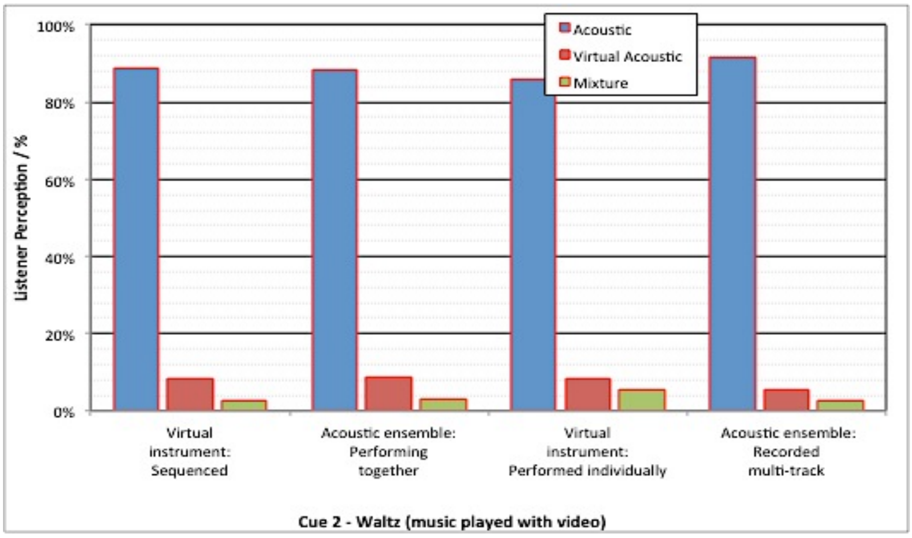

Figure 3. Plot of the summary percentage data from Table 31 for Cue 2 Waltz (music played with video).

Table 31 shows the listener responses to the four different version of Cue 2 Waltz **(QUESTION 1),** where the music cue is played with the video material and surrounding audio effects. The played versions are indicated and the listener data concerning source identification is presented underneath each version. For each example, listeners were asked to identify the source as 'acoustic', 'virtual acoustic' or 'mixture'. **Figure 3** charts these responses graphically.

		Played Versions			
		Virtual instrument: Sequenced	Acoustic ensemble: Performing together	Virtual instrument: Performed individually	Acoustic ensemble: Recorded multi-track
Strength	Weak	0.0%	0.0%	0.0%	2.9%
	Strong	100.0%	100.0%	100.0%	97.1%
Total		100.0%	100.0%	100.0%	100.0%

Table 32. Cue 2 Waltz - Emotional strength (percentage representation).

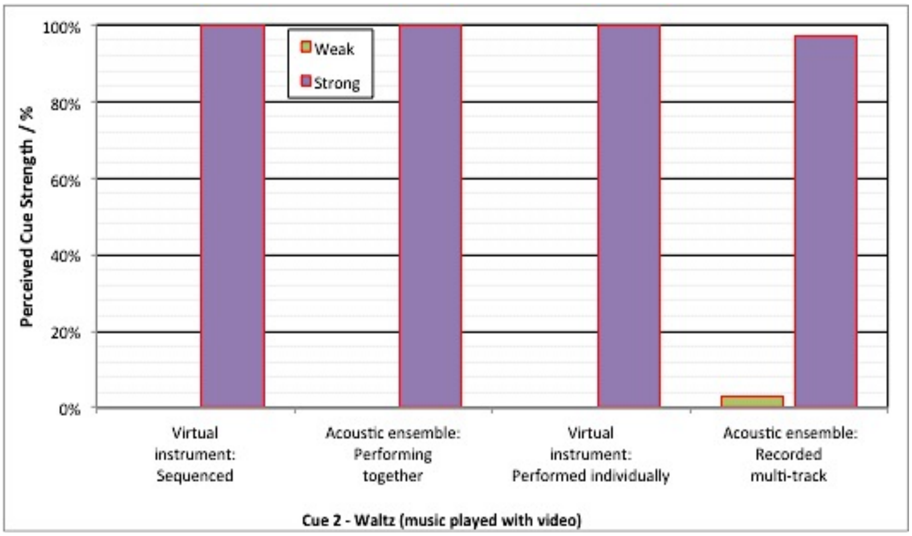

Figure 4. Cue 2 Waltz - Emotional strength (graphic representation).

Table 32 shows the listener responses to the four different version of Cue 2 Waltz **(QUESTION 3),** where the music cue is played with the video material and surrounding audio effects. The played versions are indicated and the listener data concerning source identification is presented underneath each version. For each example, listeners were asked to identify the emotional strength of the source as ‘acoustic’, ‘strong or ‘weak’. **Figure 4** charts these responses graphically.

A.1.3 Cue 1 – Superman (music played without video)

		Played Versions			
		Acoustic ensemble: Recorded multi-track	Virtual instrument: Performed individually	Virtual instrument: Sequenced	Acoustic ensemble: Performing together
Perception	Acoustic	94.4%	44.4%	41.6%	91.6%
	Virtual	5.6%	44.4%	44.4%	5.5%
	Mixture	0.0%	11.1%	13.9%	2.8%
Total		100.0%	100.0%	100.0%	100.0%

Table 33. Cue 1 Superman (music played without video) - Listener perception identification summary

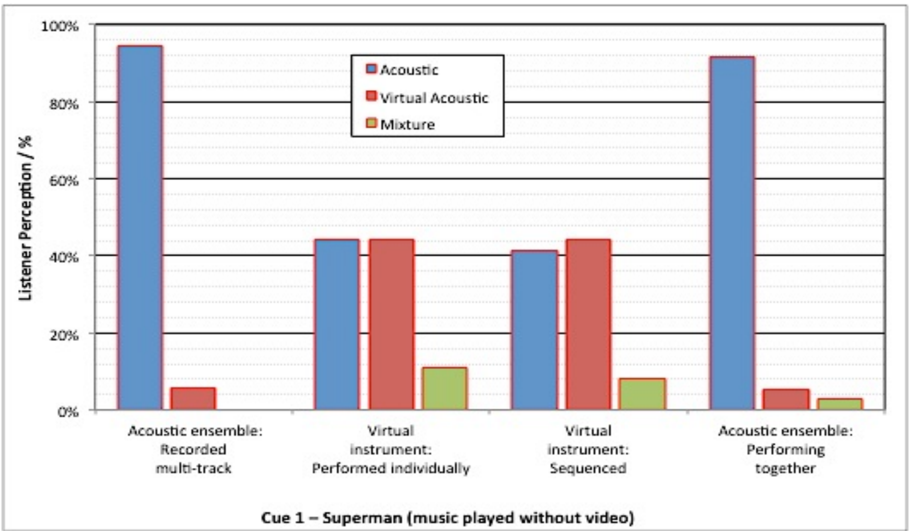

Figure 5. Plot of the summary percentage data from Table 33 for Cue 1 Superman (music played without video).

Table 33 shows the listener responses to the four different version of Cue 1 Superman **(QUESTION 1),** where the music cue is played without the video material and surrounding audio effects. The played versions are indicated and the listener data concerning source identification is presented underneath each version. For each example, listeners were asked to identify the source as ‘acoustic’, ‘virtual acoustic’ or ‘mixture’. **Figure 5** charts these responses graphically.

		Played Versions			
		Acoustic ensemble: Recorded multi-track	Virtual instrument: Performed individually	Virtual instrument: Sequenced	Acoustic ensemble: Performing together
Strength	Weak	0.0%	5.7%	2.9%	0.0%
	Strong	100.0%	94.3%	97.1%	100.0%
Total		100.0%	100.0%	100.0%	100.0%

Table 34. Cue 1 Superman - Emotional strength (percentage representation).

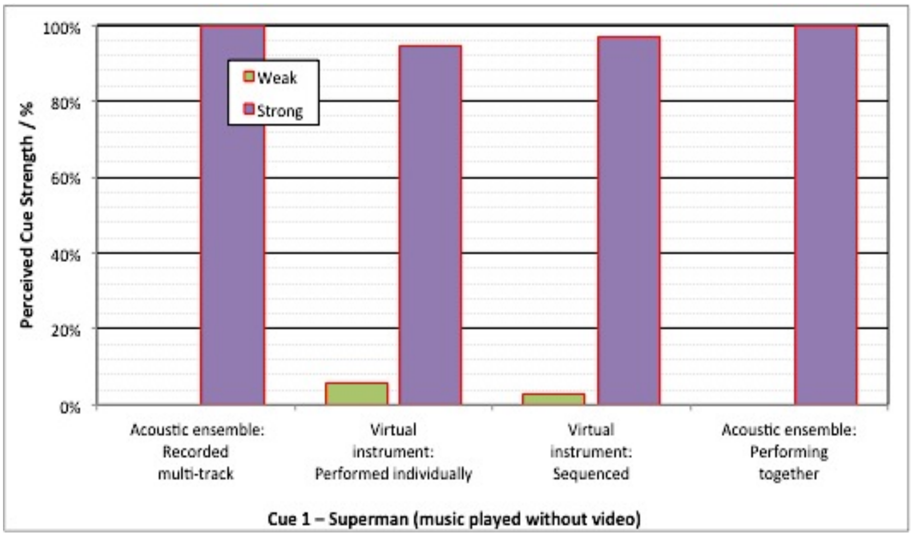

Figure 6. Cue 1 Superman - Emotional strength (graphic representation).

Table 34 shows the listener responses to the four different version of Cue 1 Superman **(QUESTION 3),** where the music cue is played without the video material and surrounding audio effects. The played versions are indicated and the listener data concerning source identification is presented underneath each version. For each example, listeners were asked to identify the emotional strength of the source as 'acoustic', 'strong or 'weak'. **Figure 6** charts these responses graphically.

A.1.4 Cue 2 - Waltz (music played without video)

		Played Versions			
		Acoustic ensemble: Recorded multi-track	Virtual instrument: Sequenced	Virtual instrument: Performed individually	Acoustic ensemble: Performing together
Perception	Acoustic	91.7%	41.6%	41.6%	88.6%
	Virtual	5.6%	47.2%	50.0%	2.8%
	Mixture	2.8%	11.1%	8.3%	8.6%
Total		100.0%	100.0%	100.0%	100.0%

Table 35. Cue 2 Waltz (music played without video) - Listener perception identification summary

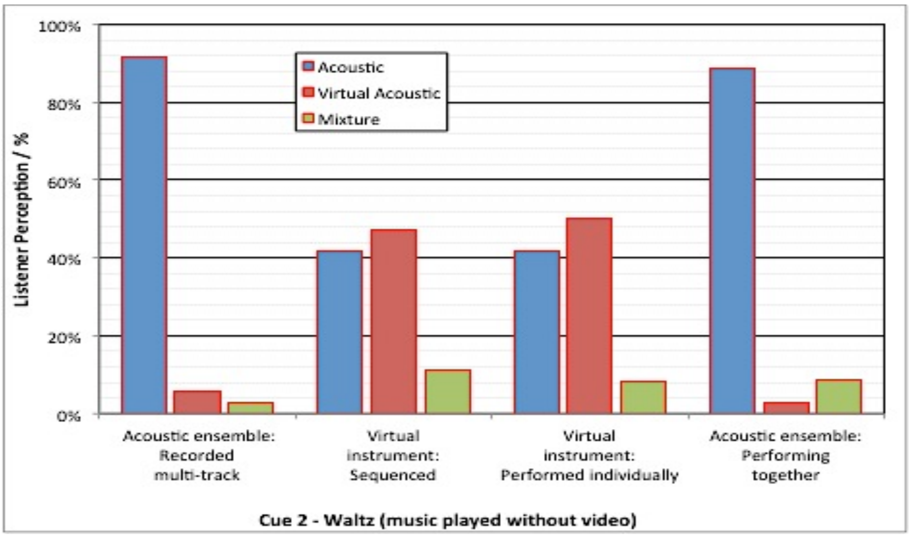

Figure 7. Plot of the summary percentage data from Table 35 for Cue 2 Waltz (music played without video).

Table 35 shows the listener responses to the four different version of Cue 2 Waltz (QUESTION 1), where the music cue is played without the video material and surrounding audio effects. The played versions are indicated and the listener data concerning source identification is presented underneath each version. For each example, listeners were asked to identify the source as 'acoustic', 'virtual acoustic' or 'mixture'. **Figure 7** charts these responses graphically.

		Played Versions			
		Acoustic ensemble: Recorded multi-track	Virtual instrument: Sequenced	Virtual instrument: Performed individually	Acoustic ensemble: Performing together
Strength	Weak	2.9%	0.0%	5.7%	2.9%
	Strong	97.1%	100.0%	94.3%	97.1%
Total		100.0%	100.0%	100.0%	100.0%

Table 36. Cue 2 Emotional strength (percentage representation).

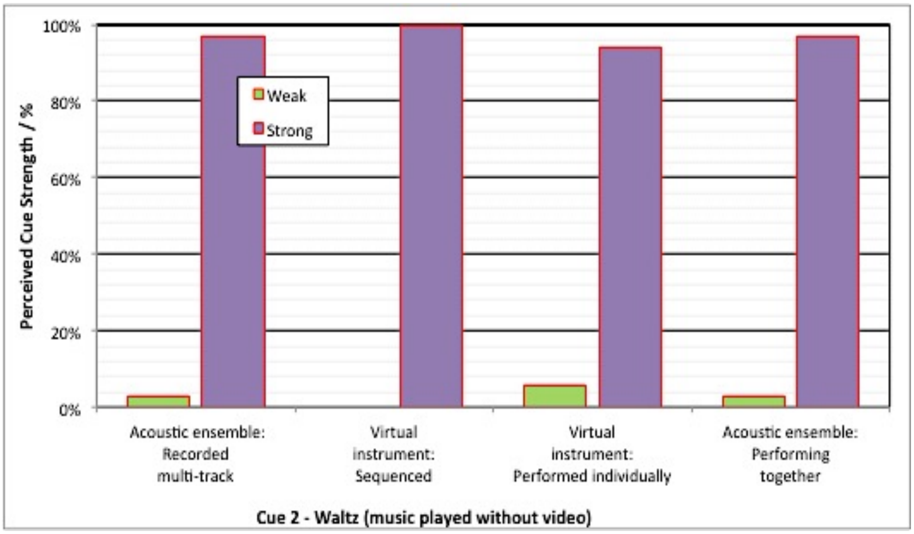

Figure 8. Emotional strength (graphic representation).

Table 36 shows the listener responses to the four different version of Cue 2 – Waltz (QUESTION 3), where the music cue is played without the video material and surrounding audio effects. The played versions are indicated and the listener data concerning source identification is presented underneath each version. For each example, listeners were asked to identify the emotional strength of the source as ‘acoustic’, ‘strong or ‘weak’. **Figure 8** charts these responses graphically.

A.1.5 Preferred example

A.1.5.1 CUE 1 – SUPERMAN (MUSIC PLAYED WITH VIDEO)

Played examples	Count	Percent
Acoustic ensemble: Performed together	10	27.7%
Acoustic ensemble: Recorded multi-track	12	33.3%
Virtual instrument: Sequenced	6	16.7%
Virtual instrument: Performed individually	8	22.2%
Total	36	100%

Table 37. Summary data of listeners' preferred example from the four Cue 1 Superman music examples (music played with video).

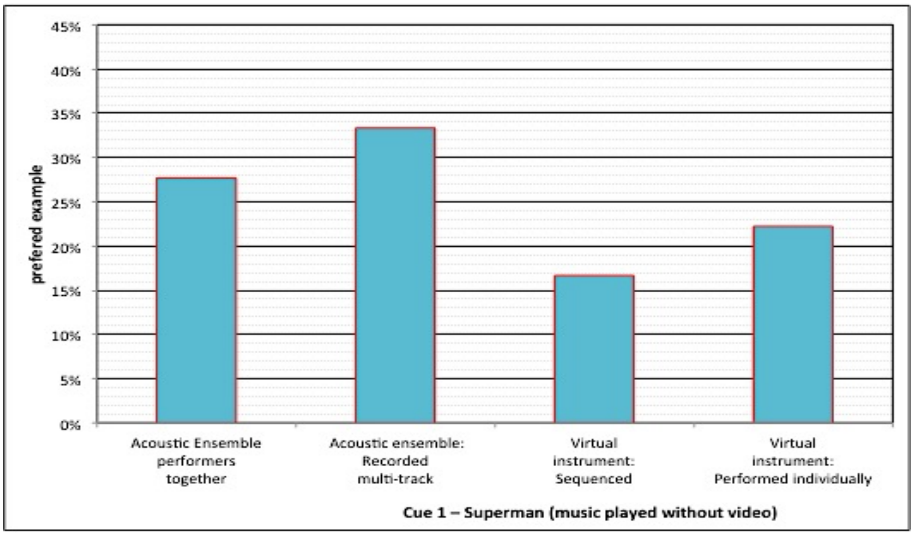

Figure 9. Preferred example (graphic representation).

Table 37 shows the listener responses to the four different version of Cue 1 Superman (QUESTION 4), where the music cue is played with the video material and surrounding audio effects. The played versions are indicated and the listener data concerning preferred example is presented next to each version. After each set of four examples, listeners were asked to nominate their preferred example. **Figure 9** charts these responses graphically.

A.1.5.2 CUE 2 – WALTZ (MUSIC PLAYED WITH VIDEO)

Played example	Count	Percent
Acoustic ensemble: Performed together	11	30.6%
Acoustic ensemble: Recorded multi-track	12	33.3%
Virtual instrument: Sequenced	9	25.0%
Virtual instrument: Performed individually	4	11.1%
Total	36	100%

Table 38. Summary data of listeners' preferred example from the four Cue 2 Waltz music examples (music played with video).

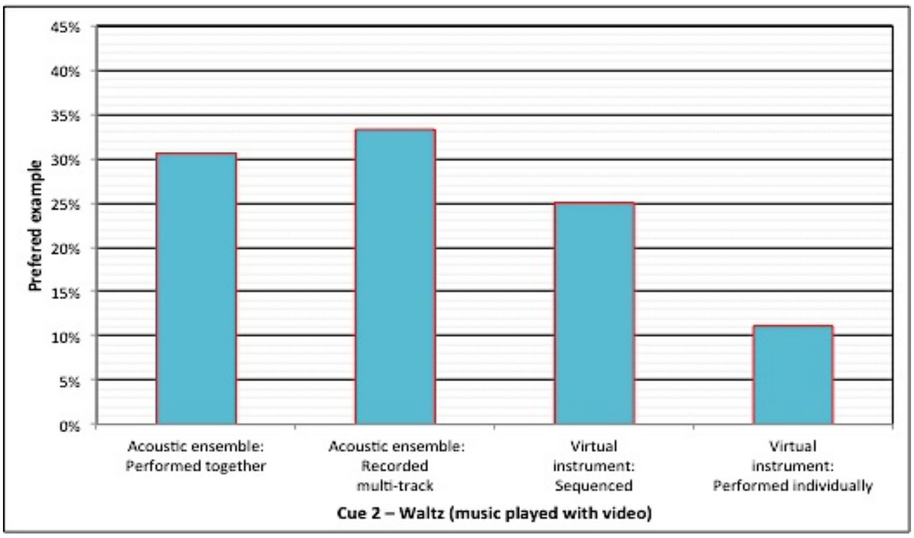

Figure 10. Preferred example (graphic representation).

Table 38 shows the listener responses to the four different version of Cue 2 Waltz (QUESTION 4), where the music cue is played with the video material and surrounding audio effects. The played versions are indicated and the listener data concerning preferred example is presented next to each version. After each set of four examples, listeners were asked to nominate their preferred example. **Figure 10** charts these responses graphically.

A.1.5.3 CUE 1 – SUPERMAN (MUSIC PLAYED WITHOUT VIDEO)

Played example	Count	Percent
Acoustic ensemble: Performed together	9	25.0%
Acoustic ensemble: Recorded multi-track	15	45.5%
Virtual instrument: Sequenced	5	13.9%
Virtual instrument: Performed individually	7	19.4%
Total	33	91.7%

Table 39. Summary data of listeners' preferred example from the four Cue 1 Superman music examples (music played without video).

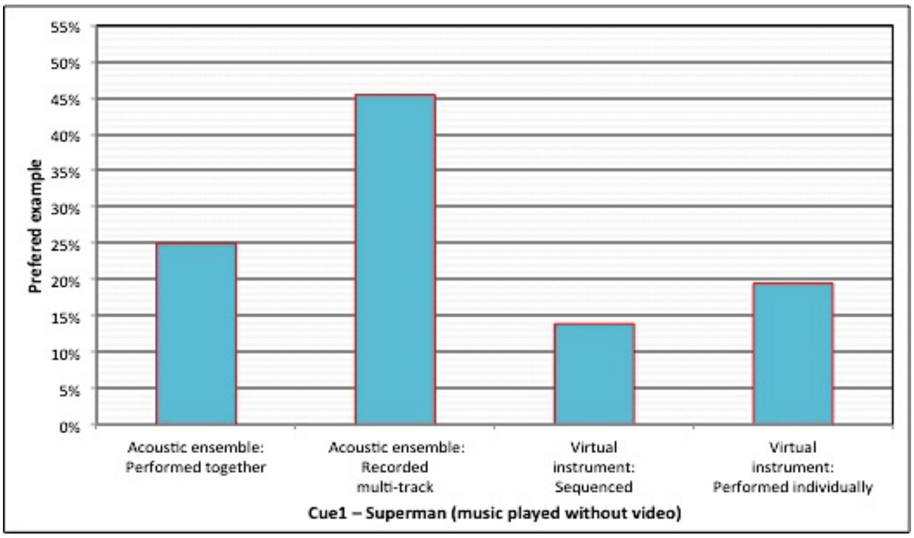

Figure 11. Preferred example (graphic representation).

Table 39 shows the listener responses to the four different version of Cue 1 Superman (**QUESTION 4**), where the music cue is played without the video material and surrounding audio effects. The played versions are indicated and the listener data concerning preferred example is presented next to each version. After each set of four examples, listeners were asked to nominate their preferred example. **Figure 11** charts these responses graphically.

A.1.5.4 Cue 2 – Waltz (music played without video)

Played examples	Count	Percent
Acoustic ensemble: Performed together	11	30.6%
Acoustic ensemble: Recorded multi-track	12	33.3%
Virtual instrument: Sequenced	7	19.4%
Virtual instrument: Performed individually	6	16.7%
Total	36	100%

Table 40. Summary data of listeners' preferred example from the four Cue 2 Waltz music examples (music played without video).

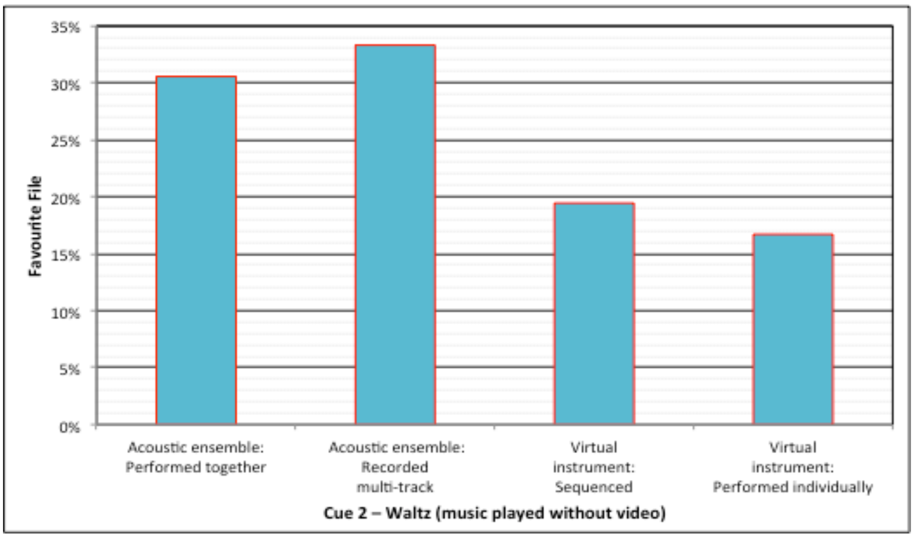

Figure 12. Preferred example (graphic representation).

Table 40 shows the listener responses to the four different version of Cue 2 Waltz (**QUESTION 4)**, where the music cue is played without the video material and surrounding audio effects. The played versions are indicated and the listener data concerning preferred example is presented next to each version. After each set of four examples, listeners were asked to nominate their preferred example. **Figure 12** charts these responses graphically.

A.1.6 Musical background of the respondents

A.1.6.1 ABILITY TO PLAY MUSICAL INSTRUMENT

Answer options	Response Percent	Response Count
Yes	47.2%	17
No	52.8%	19

Table 41. Summary data of listeners' responses.

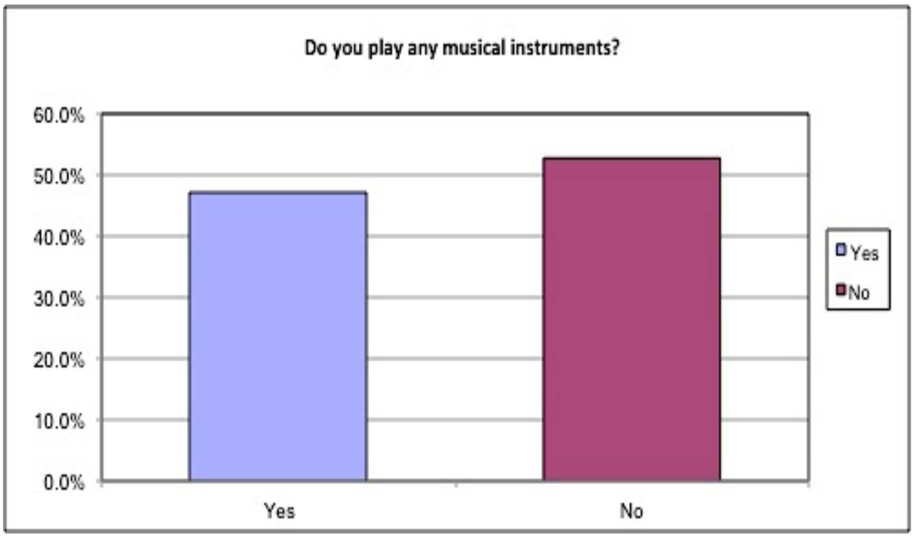

Figure 13. Ability to play musical instrument (graphic representation).

Table 41 shows the listener responses to **QUESTION 5**. After all 16 examples were played the participants were asked if they play any musical instrument. **Figure 13** charts these responses graphically.

A.1.6.2 FORMAL MUSICAL INSTRUCTION

Answer options	Response Percent	Response Count
Yes	47.2%	17
No	52.8%	19

Table 42. Summary data of listeners' responses.

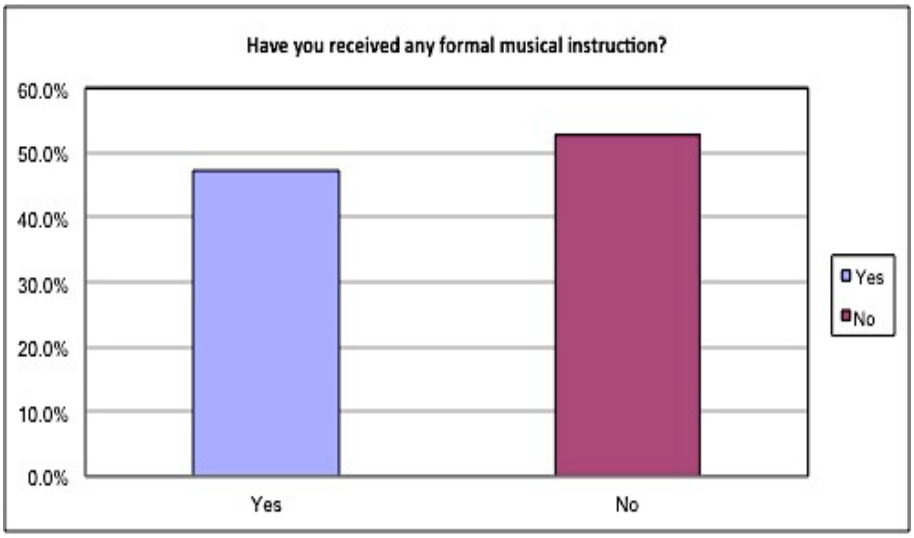

Figure 14. Formal musical instruction (graphic representation).

Table 42 shows the listener responses to **QUESTION 6**. After all 16 examples were played the participants were asked if they have received any formal musical instruction. **Figure 14** charts these responses graphically.

A.1.6.3 Level of musical proficiency

Answer Options	Response Percent	Response Count
Professional	5.8%	1
Keen amateur	94.1%	16
Hobbyist	0.0%	0

Table 43. Summary data of listeners' responses.

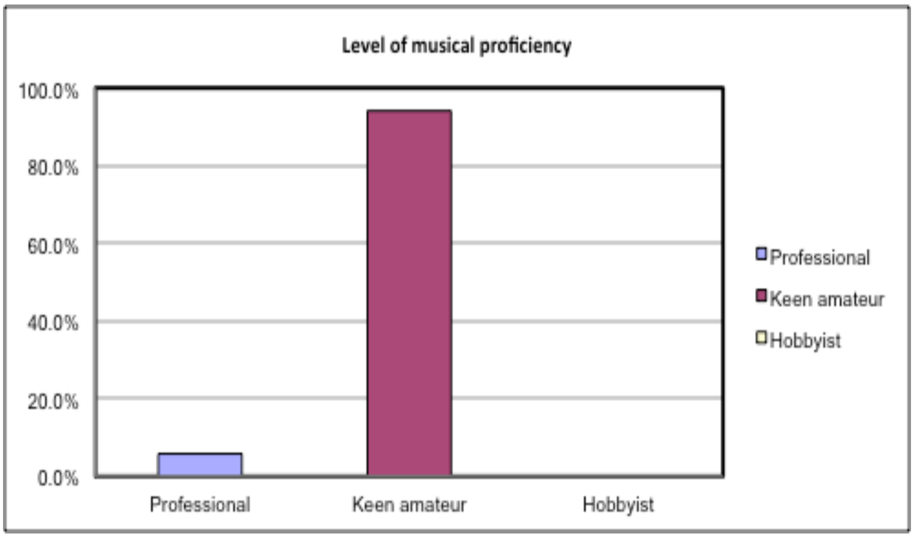

Figure 15. Level of musical proficiency (graphic representation).

Table 43 shows the listener responses to **QUESTION 7**. After all 16 examples were played the participants were asked to describe their level of musical proficiency as 'professional', 'keen amateur or 'hobbyist'. **Figure 15** charts these responses graphically.

A.1.7 Comparing Audio with Video

A.1.7.1 Cue 1 – Superman

		Played examples			
		Virtual instrument: Performed individually	Acoustic ensemble: Recorded multi-track	Virtual instrument: Sequenced	Acoustic ensemble: Performing together
Audio format	Acoustic	44.4%	94.4%	41.6%	91.6%
	Virtual	44.4%	5.6%	44.4%	5.5%
	Mixture	11.1%	0.0%	13.9%	2.8%
Video format	Acoustic	86.1%	88.9%	83.3%	88.9%
	Virtual	11.1%	8.3%	11.1%	11.1%
	Mixture	2.8%	2.8%	5.6%	0.0%
Total		100.0%	100.0%	100.0%	100.0%

Table 44. Summary percentage data of listeners' perceptions of the eight different Cue 1 Superman music files (four of which are in Video format, and four in Audio format).

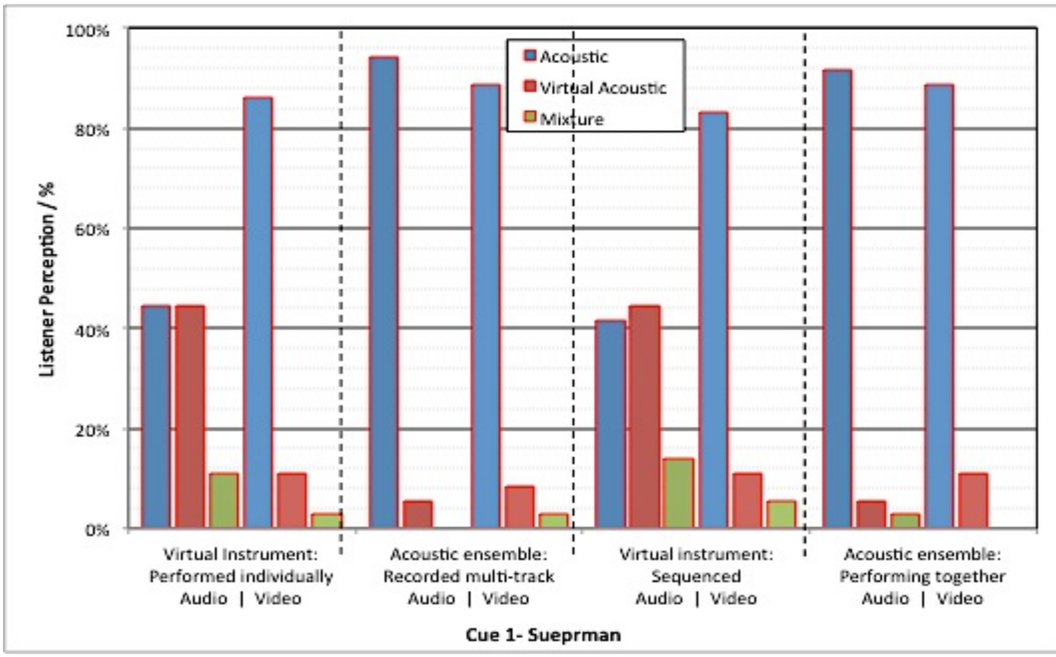

Figure 16. Plot of the summary percentage data for Cue 1 Superman music files from Table 44.

Table 44 shows the listener responses to the eight different Cue 1 Superman music files (four of which are in Video format, and four in Audio format), (**QUESTION 1**). The played versions are indicated and the listener data concerning source identification is presented underneath each version. For each example, listeners were asked to identify the source as 'acoustic', 'virtual acoustic' or 'mixture'. **Figure 16** charts these responses graphically.

A.1.7.2 Cue 2 – Waltz

		Played examples			
		Virtual instrument: Performed individually	Acoustic ensemble: Recorded multi-track	Virtual instrument: Sequenced	Acoustic ensemble: Performing together
Audio format	**Acoustic**	41.6%	91.7%	41.6%	88.6%
	Virtual	50.0%	5.6%	47.2%	8.6%
	Mixture	8.3%	2.8%	11.1%	2.9%
Video format	**Acoustic**	86.1%	91.7%	88.9%	88.6%
	Virtual	8.3%	5.5%	8.3%	2.8%
	Mixture	5.6%	2.8%	2.8%	8.6%
Total		100.0%	100.0%	100.0%	100.0%

Table 45. Summary percentage data of listeners' perceptions of the eight different Cue 2 Waltz music files (four of which are in Video format, and four in Audio format).

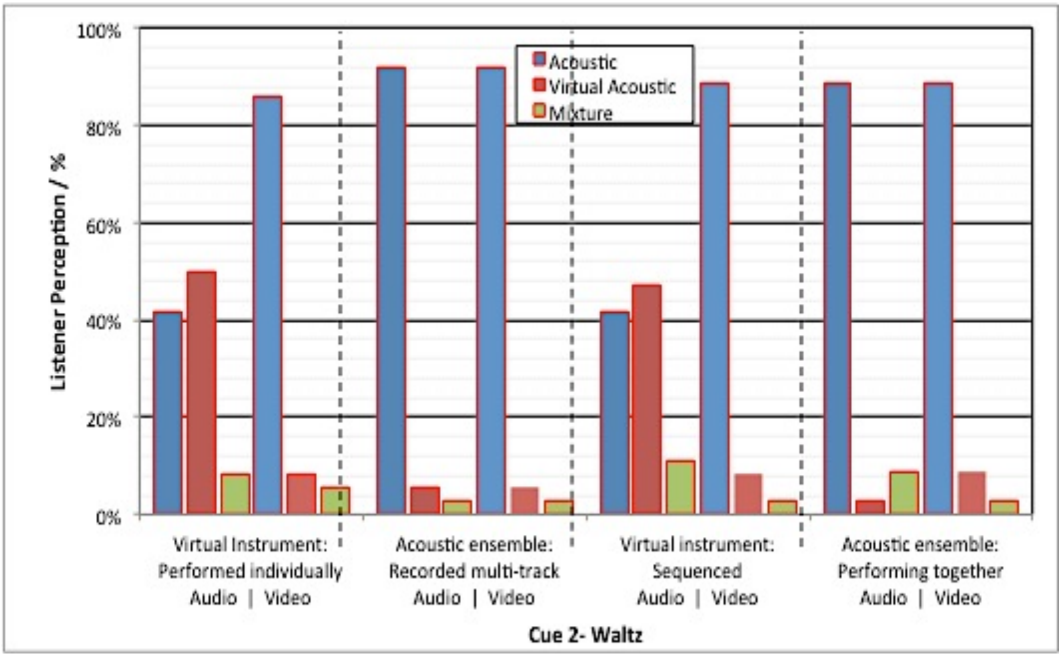

Figure 17. Plot of the summary percentage data for Cue 2 Waltz music files from Table 45.

Table 45 shows the listener responses to the eight different Cue 2 Waltz music files (four of which are in Video format, and four in Audio format), (**QUESTION 1**). The played versions are indicated and the listener data concerning source identification is presented underneath each version. For each example, listeners were asked to identify the source as 'acoustic', 'virtual acoustic' or 'mixture'. **Figure 17** charts these responses graphically.

A.1.8 Comparing Perception of Listeners who Play/Do Not Play a Musical Instrument

A.1.8.1 Cue 1 – Superman (music played with video)

	Played example							
	Virtual instrument: Performed individually		Acoustic ensemble: Recorded multi-track		Virtual instrument: Sequenced		Acoustic ensemble: Performing together	
Ability to play musical instrument	Do not play	Play	Do not play	Play	Do not play	Play	Do not play	Play
Acoustic	84.2%	88.2%	84.2%	94.1%	84.2%	82.3%	89.5%	88.2%
Virtual	15.7%	5.9%	10.5%	5.9%	10.5%	11.7%	10.5%	11.8%
Mixture	0.0%	5.9%	5.3%	0.0%	5.3%	5.9%	0.0%	0.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 46. Summary data of listeners' perceptions of four distinct Cue 1 Superman Files. The results are resolved according to whether the listener plays or does not play a musical instrument.

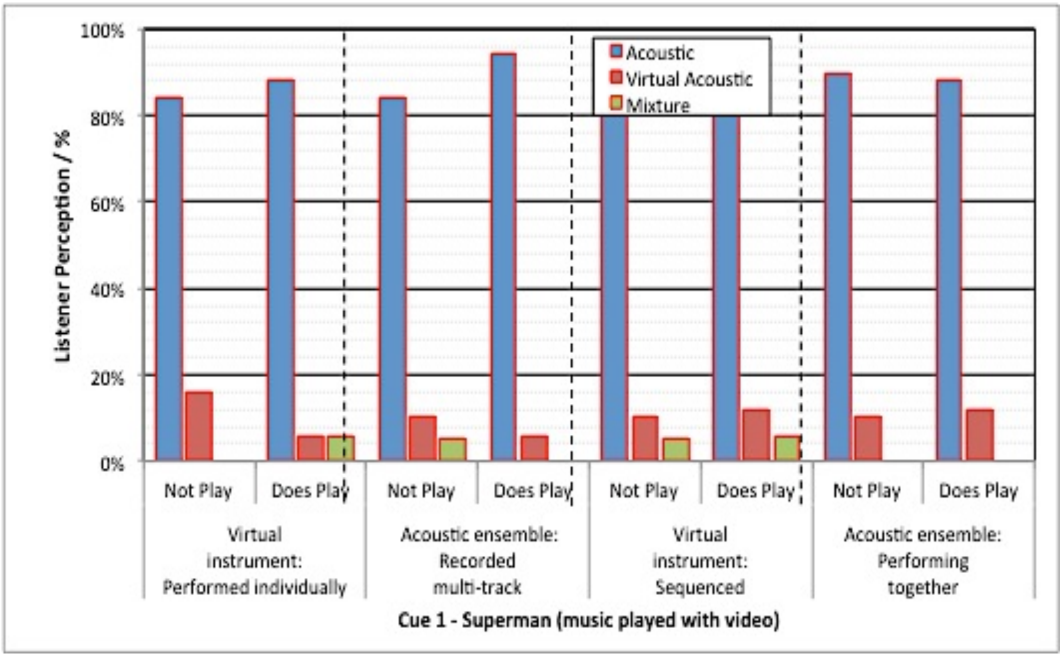

Figure 18. Plot of the summary percentage data for Cue 1 Superman music files from Table 46.

Table 46 shows the listener responses to the four different versions of Cue 1 Superman (Video, Music, FX) according to whether the listener 'Plays' or does 'Does not play' a musical instrument in such a way as to facilitate comparison. The played versions are indicated horizontally and the source identification is presented underneath each version for the two demographic groups. For each example, listeners were asked to identify the source as 'acoustic', 'virtual acoustic' or 'mixture'. **Figure 18** charts these responses graphically.

A.1.8.2 CUE 2 – WALTZ (MUSIC PLAYED WITH VIDEO)

	Played example							
	Virtual instrument: Performed individually		Acoustic ensemble: Recorded multi-track		Virtual instrument: Sequenced		Acoustic ensemble: Performing together	
Ability to play musical instrument	Not Play	Does Play	Not Play	Does Play	Not Play	Does Play	Not Play	Does Play
Acoustic	84.2%	94.1%	89.5%	88.2%	84.2%	82.4%	89.5%	94.1%
Virtual	10.5%	5.9%	10.5%	5.9%	15.8%	5.8%	5.3%	5.9%
Mixture	5.3%	0.0%	0.0%	5.9%	0.0%	11.8%	5.3%	0.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 47. Summary data of listeners' perceptions of four distinct Cue 2 Waltz Files. The results are resolved according to whether the listener plays or does not play a musical instrument.

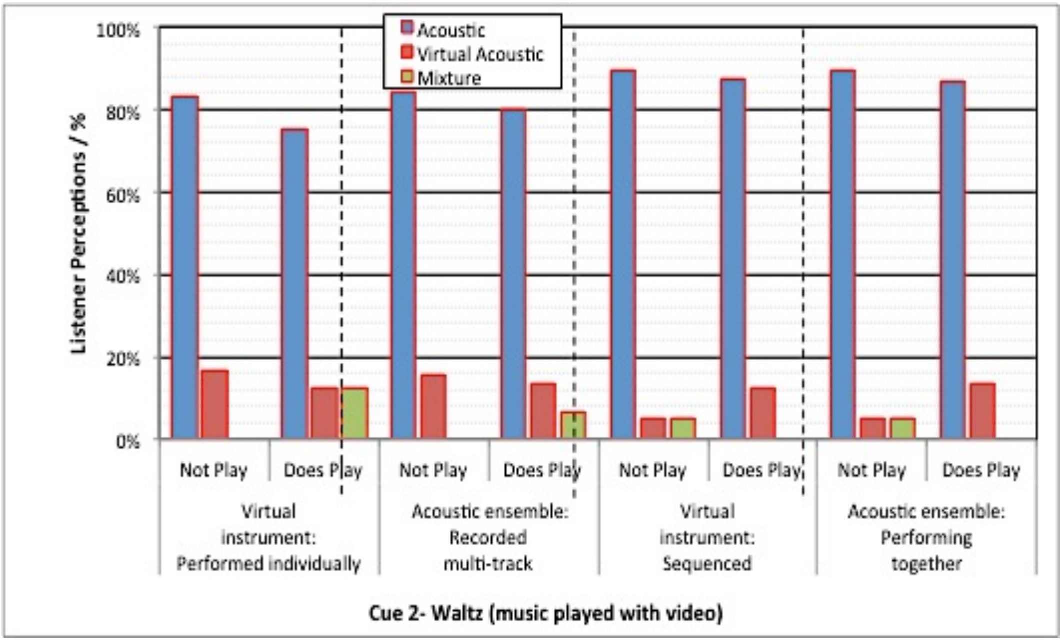

Figure 19. Plot of the summary percentage data for Cue 2 Waltz music files from Table 47.

Table 47 shows the listener responses to the four different versions of Cue 2 Waltz (Video, Music, FX) according to whether the listener 'Plays' or does 'Does not play' a musical instrument in such a way as to facilitate comparison. The played versions are indicated horizontally and the source identification is presented underneath each version for the two demographic groups. For each example, listeners were asked to identify the source as 'acoustic', 'virtual acoustic' or 'mixture'. **Figure 18** charts these responses graphically.

A.1.9.3 CUE 1 – SUPERMAN (MUSIC PLAYED WITHOUT VIDEO)

	Played example							
	Virtual instrument: Performed individually		Acoustic ensemble: Recorded multi-track		Virtual instrument: Sequenced		Acoustic ensemble: Performing together	
Ability to play musical instrument	Not Play	Does Play	Not Play	Does Play	Not Play	Does Play	Not Play	Does Play
Acoustic	63.1%	23.5%	89.5%	100.0%	63.2%	20.0%	89.5%	94.1%
Virtual	26.3%	64.7%	10.5%	0.0%	31.6%	66.7%	5.3%	5.9%
Mixture	10.5%	11.8%	0.0%	0.0%	5.3%	13.3%	5.3%	0.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 48. Summary data of listeners' perceptions of four distinct Cue 1 Superman Files. The results are resolved according to whether the listener plays or does not play a musical instrument.

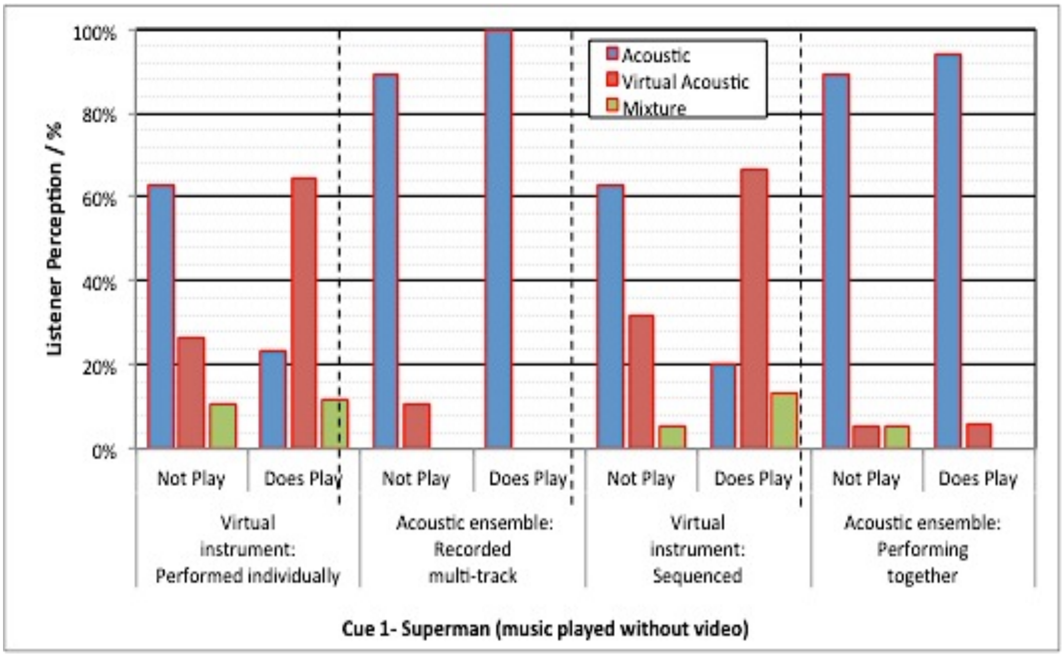

Figure 20. Plot of the summary percentage data for Cue 1 Superman music files from Table 48.

Table 48 shows the listener responses to the four different versions of Cue 1 Superman (music played without video and fx) according to whether the listener 'Plays' or does 'Does not play' a musical instrument in such a way as to facilitate comparison. The played versions are indicated horizontally and the source identification is presented underneath each version for the two demographic groups. For each example, listeners were asked to identify the source as 'acoustic', 'virtual acoustic' or 'mixture'. **Figure 18** charts these responses graphically.

A.1.9.4 CUE 2 – WALTZ (MUSIC PLAYED WITHOUT VIDEO)

	Played example							
	Virtual instrument: Performed individually		Acoustic ensemble: Recorded multi-track		Virtual instrument: Sequenced		Acoustic ensemble: Performing together	
Ability to play musical instrument	Not Play	Does Play	Not Play	Does Play	Not Play	Does Play	Not Play	Does Play
Acoustic	52.6%	29.4%	94.7%	88.2%	57.8%	23.5%	84.2%	94.1%
Virtual	36.8%	64.7%	5.3%	5.9%	21.1%	76.5%	10.5%	5.9%
Mixture	10.5%	5.9%	0.0%	5.9%	21.1%	0.0%	5.3%	0.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 49. Summary data of listeners' perceptions of four distinct Cue 2 Waltz Files. The results are resolved according to whether the listener plays or does not play a musical instrument.

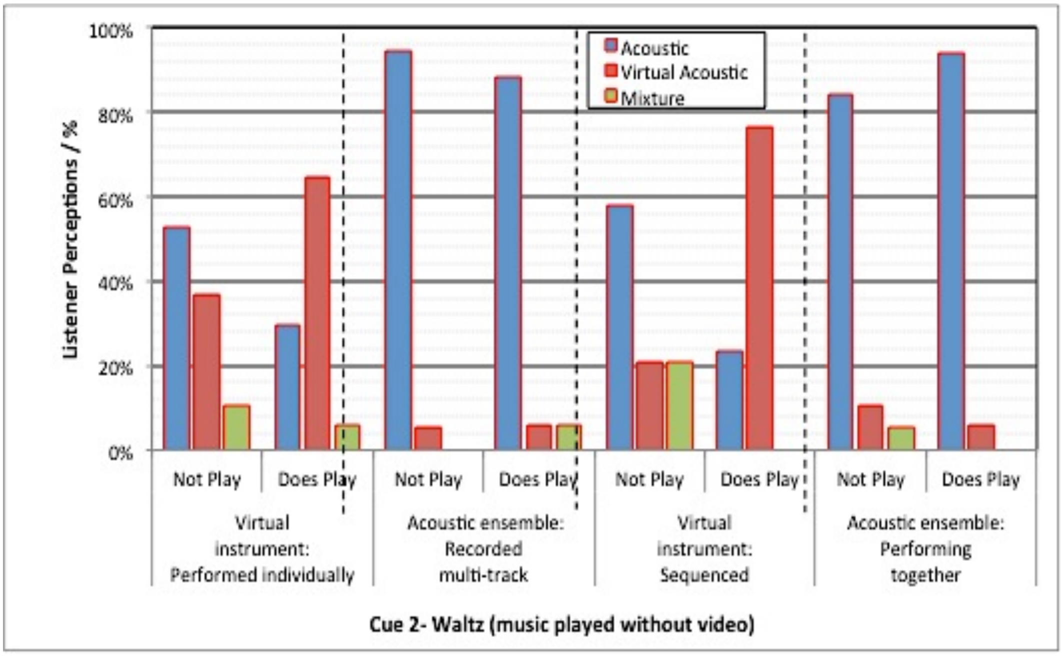

Figure 21. Plot of the summary percentage data for Cue 2 Waltz music files from Table 49.

Table 49 shows the listener responses to the four different versions of Cue 2 Waltz (Video, Music, FX) according to whether the listener 'Plays' or does 'Does not play' a musical instrument in such a way as to facilitate comparison. The played versions are indicated horizontally and the source identification is presented underneath each version for the two demographic groups. For each example, listeners were asked to identify the source as 'acoustic', 'virtual acoustic' or 'mixture'. **Figure 18** charts these responses graphically.

A.2 Notation of the musical examples

SUPERMAN

SUPERMAN

SUPERMAN


Violin II
= 100
1
Vln. II
5
= 100
Vln. II
9
Vln. II
13
Vln. II
17
Vln. II
21
Vln. II
25
Vln. II
29
Vln. II
33

WALTZ

WALTZ


Viola
= 92
1
Vla.
5
Vla.
9
Vla.
13
Vla.
17
Vla.
21
Vla.
25
Vla.
29
Vla.
33
Vla.
37
Vla.
41
Vla.
45

WALTZ

WALTZ

Supplementary material

DVD 1: Audio files

- Superman Acoustic ensemble performing together.wav
- Superman Acoustic ensemble recorded multi-track.wav
- Superman Virtual instruments performed individually.wav
- Superman Virtual instruments sequenced.wav
- Waltz Acoustic ensemble performing together.wav
- Waltz Acoustic ensemble recorded multi-track.wav
- Waltz Virtual instruments performed individually.wav
- Waltz Virtual instruments sequenced.wav

DVD 2: Video files

- Superman Acoustic ensemble performing together.mpeg
- Superman Acoustic ensemble recorded multi-track.mpeg
- Superman Virtual instruments performed individually.mpeg
- Superman Virtual instruments sequenced.mpeg
- Waltz Acoustic ensemble performing together.mpeg
- Waltz Acoustic ensemble recorded multi-track.mpeg
- Waltz Virtual instruments performed individually.mpeg
- Waltz Virtual instruments sequenced.mpeg

USB: Audio and Video files

Audio files

- Superman Acoustic ensemble performing together.wav
- Superman Acoustic ensemble recorded multi-track.wav
- Superman Virtual instruments performed individually.wav
- Superman Virtual instruments sequenced.wav
- Waltz Acoustic ensemble performing together.wav
- Waltz Acoustic ensemble recorded multi-track.wav
- Waltz Virtual instruments performed individually.wav
- Waltz Virtual instruments sequenced.wav

Video files

- Superman Acoustic ensemble performing together.mpeg
- Superman Acoustic ensemble recorded multi-track.mpeg
- Superman Virtual instruments performed individually.mpeg
- Superman Virtual instruments sequenced.mpeg
- Waltz Acoustic ensemble performing together.mpeg
- Waltz Acoustic ensemble recorded multi-track.mpeg
- Waltz Virtual instruments performed individually.mpeg
- Waltz Virtual instruments sequenced.mpeg

Bibliography

Ariza, C. (2009). The interrogator as critic. *Computer Music Journal , 33* (2), 48-70.

Asher, J. (2010 5-August). *Tech Talk With Composer Bruce Miller*. Retrieved 2014 28-September from Film Music Magazine: http://www.filmmusicmag.com/?p=6020

Audio Media. (1996 January). *Fairlight: the whole story*. Retrieved 2014 02-September from A Nerd: http://anerd.com/fairlight/fairlightstory.htm

Avanzinia, F., Banka, B., Borina, G., De Polia, G., Fontanac, F., & Rocchessoc, D. (N.D.). *Musical instrument modeling: the case of the piano.* Padova: University of Padova.

Awde, N., Bradley, J., Smith, M., Unwin, G., Pinder, M., McDonald, I., et al. (2008). *Mellotron: The Machine and the Musicians that Revolutionised Rock.* Desert Hearts.

Bates, J. (1994 01-July). The Role of Emotion: in believable agents. *Communications of the ACM , 37* (7), pp. 122-125.

Bazzana, K. (2013). *Biography.* Retrieved 2014 24-August from Glenn Gould: http://www.glenngould.com/us/biography

Beal, J. (2014). *Effects and multimedia*. Retrieved 2014 15-September from EastWest Sounds Online: http://www.soundsonline.com/Effects-and-Multimedia

Bech, S., & Zacharov, N. (2007). *Perceptual Audio Evaluation – Theory, Method and Application.* John Wiley & Sons.

Bernardini, N., & De Poli, G. (2007). The Sound and Music Computing Field: Present and Future. *Journal of New Music Research , 36* (3), 143-148.

Blom Pedersen, L. (2013). *A Study in Perceived Believability: Utilizing Visual Movement to alter the Level of Detail.* Copenhagen: Aalborg University .

Boley, J., & Lester, M. (2009). Statistical Analysis of ABX Results Using Signal Detection Theory. *Audio Engineering Society Convention, 127*.

Braun, H. J. (2004). Book Review: Modern Sounds. *Social Studies of Science, 34* (5), 809-817.

Brown, A. (2001). How the computer assists composers: A survey of contemporary practise. *Australasian Computer Music Conference* (pp. 9-16). Sydney: University of Western Sydney.

Burnett, M. (2008). *Can You Hear the Difference? A sonic comparison between specific Digidesign Audio Interfaces*.

Burt, G. (1995). *The Art of Film Music.* Boston: Northeastern University Press.

Candy, L. (2006). *Practice based research: A guide.* Creativity and Cognition Studios. Sydney: University of Technology.

Chion, M. (1994). *Audio-vision: sound on screen.* New York: Columbia University Press.

Collins, N. (2008). Infno: Generating Synth Pop and Electronic Dance Music on Demand. *Proceedings of the International Computer Music Conference.*

Collins, N. (2006). *Towards autonomous agents for live computer music: Real-time machine listening and interactive music systems.* Cambridge University.

Cook, P. (2001). Principles for designing computer music controllers. *Conference on New Interfaces for Musical Expression.* Seattle.

Davies, H. (1996). A History of Sampling. *Organised Sound, 1* (1), 3-11.

Dillon, S. C. (2007). Examining meaningful engagement: Musicology and virtual music making environments. *Conference of the Australian and New Zealand Musicological Societies* (pp. 1-17). Brisbane: Islands.

Dobson, M. C., & Gaunt, H. F. (2013). Musical and social communication in expert orchestral performance . *Psychology of Music* , 1-19.

Dolhansky, B., McPherson, A., & Kim, Y. E. (2011). *Designing an Expressive Virtual Percussion Instrument.* Retrieved 2014 15-January from SMC Network: http://server.smcnetwork.org/system/files/smc2011_submission_71.pdf

Draper, P. (2008). Music two-point-zero: music, technology and digital independence. *Journal of Music, Technology and Education, 1* (2), 137-152.

Duignan, M., Noble, J., & Biddle, R. (2010). Abstraction and Activity in Computer-Mediated Music Production. *Computer Music Journal, 34* (4), 22-33.

Ekeberg Henriksen, F. (2002). *Space in Electroacoustic Music: Composition, Performance and Perception of Musical Space.* Retrieved 2014 1-February from www.frankenberg.no: http://japan.srnikf.com/contact_form/navi.html?pro%20active%20acne

Ellis Geiger, R. J. (2007). Trends In Contemporary Hollywood Film Scoring:A Synthesised Approach For Hong Kong Cinema. *PhD Dissertation*. Leeds: The University Of Leeds, School of Music.

Ellis-Geoger, R. (2007). Trends in contemporary Hollywood film scoring: a synthesised approach for Hong Kong cinema. University of Leeds.

Fikentscher, K. (2007). The Producer as Composer: Shaping the Sounds of Popular Music by Virgil Moorefield. *Popular Music, 26* (3), 523-525.

Folmann, T. (2014 February). *Interview with Blake Neely*. Retrieved 2014 28-September from 8DIO: http://8dio.com/blog/blakeneely/

Gabrielli, L., Squartini, S., & Välimäki, V. (2011, October). A subjective validation method for musical instrument emulation. In Audio Engineering Society Convention 131. Audio Engineering Society.

Gilreath, P. (2006). *The Guide to MIDI Orchestration.* Atlanta: Music Works.

Goebl, W., Dixon, S., De Poli, G., Friberg, A., Bresin, R., & Widmer, G. (2008). 'Sense' in Expressive Music Performance: Data Acquisition, Computational Studies, and Models. In M. Leman, D. Cirotteau, M. Leman, & D. Cirotteau (Eds.), *Sound to Sense, Sense to Sound* (pp. 1-36). Berlin: Logos.

Gonski, R. (2009). Symphonic mind: States of consciousness in orchestral performance. *Contemporary Music Review, 14* (3), 55-64.

Goodwin, A. (1988). Sample and hold: pop music in the digital age of reproduction. *Critical Quarterly*, 34-49.

Grachten, M., & Widmer, G. (2011). *Explaining musical expression as a mixture of basis functions.* Retrieved 2015 January 17 from SMC Network: http://smcnetwork.org/smc_papers-2011052

Gratier, M. (2008). Grounding in musical interaction: Evidence from jazz performances. *Musicae Scientiae* , 71-110.

Gray, C. (1996). *Inquiry through Practice: Developing appropriate research strategies.* Retrieved 2014 8-September from http://www2.rgu.ac.uk/criad/cgpapers/ngnm/ngnm.htm

Gurevich, M., & Fyans, A. C. (N.D.). *Digital Musical Interactions: Performer-System Relationships and their Perception by Spectators.* Queens University Belfast. Belfast: Sonic Arts Research Centre.

Hamm, S. (2010). *Computers and Live Performance: Interactive, or Interference?* The University of Florida, Florida Electroacoustic Music Studio. Florida: Sam Hamm.

Hennick, M., Hutter, I., & Bailey, A. (2010). *Qualitative Research Methods.* Sage.

Hickman, R. (2006). *Reel Music.* New York: Norton.

Hjortkjær, J., & Nielbo, F. (2010). *A Perceptual study on dynamical form in music.* Retrieved 2014 January 15 from SMC Network: http://smcnetwork.org/files/proceedings/2010/61.pdf

Holland, S., Wilkie, K., Mulholland, P., & Seago, A. (2013). *Music and Human-Computer Interaction.* (S. Holland, K. Wilkie, P. Mulholland, & A. Seago, Eds.) London: Springer.

Isham, M. (2010 11-November). *Hollywood composer Mark Isham talks film scores and virtual instruments*. Retrieved 2014 28-September from Time and Space: http://www.timespace.com/features/1814/hollywood-composer-mark-isham-talks-film-scores-and-virtual-instruments.html

Johnston, A., Candy, L., & Edmonds, E. (2008). *Designing and evaluating virtual musical instruments: facilitating conversational user interaction.* Creativity and Cognition Studios. Sydney: Elsevier.

Juslin, P. N., Karlsson, J., Lindstrom, E., Friberg, A., & Schoonderwaldt, E. (2006). Play It Again With Feeling: Computer Feedback in Musical Communication of Emotions. *Journal of Experimental Psychology: Applied, 12* (2), 79-95.

Karlin, F., & Wright, R. (1990). *On the Track: A Guide to Contemporary Film Scoring.* New York: Schrimer Books.

Karray, F., Alemzadeh, M., Saleh, J. A., & Nours Arab, M. (2008). Human-Computer Interaction: Overview on State of the Art. *International Journal on Smart Sensing and Intelligent Systems, 1* (1), 138-159.

Kerins, M. (2011). *Beyond Dolby: Cinema in the Digital Sound Age.* Bloomingdale: Indiana University Press.

Kim, H., Di Giacomo, T., Egges, A., Lyard, E., Garchery, S., & Magnenat Thalmann, N. (2004). *Believable Virtual Environment: Sensory and Perceptual Believability.* Geneva: MIRALab, University of Geneva.

Kingwell, M. (2009). *Extraordinary Canadians: Glenn Gould.* Toronto: Penguin Canada.

Kirke, A., & Miranda, E. R. (2013). *Guide to Computing for Expressive Music Performance.* (A. Kirke, & E. R. Miranda, Eds.) London: Springer.

Leete, N. (1999 April). *Fairlight Computer*. Retrieved 2014 2-September from Sound on Sound: http://www.soundonsound.com/sos/apr99/articles/fairlight.htm

Lipscomb, S., & Zehnder, S. (2004). *Immersion in the virtual environment: The effect of a musical score on the video gaming experience.* Kyushu University, School of Music. Evanston: Northwestern University.

Love, M. (2013). Collective Conceptualization - A Methodology for Contemporary Film Score Composers: A Portfolio of Compositions (disposed as a film score) and Exegesis. University of Adelaide.

Luciani, A., Florens, J. L., Courousse, D., & Castet, J. (2009). Ergotic Sounds: A New Way to Improve Playability, Believability and Presence of Virtual Musical Instruments. *Journal of New Music Research, 38* (3), 309-323.

Machover, T. (N.D.). *Hyperinstruments*. Retrieved 2014 2-September from Opera of the Future: http://opera.media.mit.edu/projects/hyperinstruments.html

Mack, N., Woodsong, C., MacQueen, K., Guest, G., & Namey, E. (2005). *Qualitative Research methods: a data collector's field guide.* North Carolina: Family Health International.

Magnenat Thalmann, N., Kim, H., Egges, A., & Garchery, S. (2005). Believability and Interaction in Virtual Worlds. *International Multi-Media Modelling Conference* (pp. 2-9). IEEE Computer Society Press.

Marsella, S., & Gratch, J. (2002). A step toward irrationality: using emotion to change belief. *First International Joint Conference on Autonomous Agents and Multiagent Systems* (pp. 1-8). Bologna: AAMAS.

McGuire, S., & Pritts, R. (2008). *Audio Sampling: a Practical Guide.* Focal Press.

McPherson, A., & Kim, Y. (2011). Multidimensional gesture sensing at the piano. *Proceedings of the 29th ACM conference on human factors in computing systems.* Vancouver.

Meehan, P. (1999 March). *Creating Orchestral Music With Sampled Instruments*. Retrieved 2014 28-September from Sound on Sound: http://www.soundonsound.com/sos/mar99/articles/orchestra.htm

Murch, W. (1994). Foreword. In M. Chion, *Audio-Vision* (pp. vii-xxiiiv). New York: Columbia University Press.

Nash, C. (2011). *Supporting virtuosity and flow in computer music.* University of Cambridge. Cambridge: Chris Nash.

O'Mahony, M., & Rousseau, B. (2003). Discrimination Testing: a few ideas. *Food Quality and Preference, 14* (2), 157-164.

Palm, W. (2008 30-November). Vintage Synthesizers. *Keyboard Magazine* .

Pearce, M., & Wiggings, G. (2001). Towards a Framework for the Evaluations of Machine Compositions. *Symposium on Artifical Intelligence and Creativity in the Arts and Sciences.* Brighton: SSAISB.

Pedersen, J., & Jorgensen, T. (2005). *Localization performance of real and virtual sound sources.* Copenhagen: Aalborg University.

Pedersen, L. (2013). *A Study in Perceived Believability.* Copenhagen: Aalborg University.

Pejrolo, A., & DeRosa, R. (2009). Acoustic and midi orchestration for the contemporary composer. CRC Press.

Platz, F., & Reinhard, K. (2012). When the eye listens: a meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Perception, 30* (1), 71-83.

Ramshaw, P. (2006). Is Music Production Now a Composition Process? *Paper presented to the First Annual Conferences on the Art of Record Production.* London.

Robjohns, H. (1999 April). *Recording Strings*. Retrieved 2014 20-September from Sound on Sound: http://www.soundonsound.com/sos/apr99/articles/recstrings.htm

Rualark. (n.d.). *Creating sample-based orchestral arrangements*. From Alch Music Projects: http://alch.16mb.com/articles/csboa/

Russ, M. (2004). *Sound Synthesis and Sampling.* Taylor and Francis.

Russ, M. (1994 July). *Yamaha VL1*. Retrieved 2014 2-September from Sound on Sound: http://www.soundonsound.com/sos/1994_articles/jul94/yamahavl1.html

Savioja, L., Huopaniemi, J., Lokki, T., & Vaananen, R. (1999). Creating Interactive Virtual Acoustic Environments. *Journal of Audio Engineering Society, 47* (9), 675-705.

Schertenleib, S., Gutiérrez, M., Vexo, F., & Thalmann, D. (2004). *Conducting a Virtual Orchestra.* Swiss Federal Institute of Technology. Lausanne: IEEE Computer Society.

Skelton, M. (N.D.). *Paths of Fantastic Film Music: Examining Film Scoring Techniques in Films that Alter Space, Time and Death*. Retrieved 2014 28-September from Film Score Monthly: http://www.filmscoremonthly.com/features/skelton.asp

Smith, H., & Dean, R. (2009). *Practice-led research, research-led practice in the creative arts.* Edinburgh: University Press.

Smith, J. O. (2004). Virtual acoustic musical instruments: Review and update. *Journal of New Music Research, 33* (3), 283-304.

Stewart, D. (1999 December). *Creating Realistic Sample-based Orchestral Arrangements, Part 1*. Retrieved 2014 28-September from Sound on Sound: http://www.soundonsound.com/sos/dec99/articles/orchestral.htm

Sundberg, J. (1977). The Acoustics of the Singing Voice. *Scientific American*, 82-91.

Sundstrup, L. (2009). *The virtual orchestra: a systematic method of realising music composition through sample-based orchestral simulation.* Wollongong: University of Wollongong.

Thornely, S. (ND). *The Impact of Computer Music Technology on Music Production.* Retrieved 2014 15-January from Steven Thornely: Professional, Committed, Passionate: http://stevethornely.files.wordpress.com/2012/05/the-impact-of-computer-technology-on-world-music-steven-thornely.pdf

Trochidis, K., Delbe, C., & Bigand, E. (2011). *Investigation of the relationships between audio features and induced emotions in contemporary western music .* Retrieved 2014 15-January from SMC Network: http://smcnetwork.org/smc_papers-2011114

Tsay, C. J. (2013). *Sight over sound in the judgment of music performance.* University College London, Department of Management Science and Innovation. London: PNAS.

Wanderley, M. M. (2001, September). Gestural control of music. In International Workshop Human Supervision and Control in Engineering and Music (pp. 632-644).

Wang, G. (2007). A history of programming and music. In N. Collins, & L. d'Escrivian, *Electronics Music* (pp. 55-70). Cambridge University Press.

Wierzbicki, J., Platte, N., & Roust, C. (. (2012). *The Routledge Film Music Sourcebook.* Routledge.

Williamon, A., & Davidson, J. W. (2002). Exploring Co-Performer Communication. *Musicae Scientiae, 1* (1), 53-72.

Wingstedt, J., Brändström, S., & Berg, J. *Narrative Music, Visuals and Meaning in Film.* Stockholm.

Zhang, Z., & Johnston, J. (2013). *Patent No. 8,620,009.* United States.