Recording piano in surround: discovering preferences, investigating auditory imagery, and establishing physical predictors.

Kim, Sungyoung

Doctor of Philosophy

Schulich School of Music

McGill University

Montreal,Quebec

2009-02-01

–

ACKNOWLEDGEMENTS

I would like to thank my wife, Eunhee Kim, the meaning of my life. Without her sincere belief and love, I could not have started my study. I sincerely thank professor William L. Martens who has endured and guided me with excellent insight and inspiration so that I could conduct this research and bring it to fruition. Professor Wieslaw Woszczyk also deserves thanks for being a great mentor and supporting my research with his foresight in multichannel recording and reproduction. It would be impossible to finish my acknowledgements without mentioning Kent Walker who has helped, envisaged and shared a vision with me through the Masters and Ph.D. programs. Thanks also to professor Martha DeFrancisco for conducting the recording session and giving uncountable advice for the recording and music. Thanks to my parents, who believed in and supported my study, and to my two kids, Seohee and Joseph, for putting up with their father who spent more time in the laboratory and library instead of being with them. More than any thanks, I thank God who leads my way always. For their financial support, I thank *Fonds Québécois de la Recherche sur la Société et la Culture* (FQRSC), *Valorisation-Recherche Québec* (VRQ) of the Government of Quebec within the project *Real-time Communication of High-resolution Multi-sensory Content via Broadband Networks* and Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT). Thanks also to the Yamaha corporation for providing support the preparation of this text.

ABSTRACT

In order to develop a better understanding of the differences in the multichannel reproduction of solo piano music that result from the use of various popular multichannel microphone techniques, an experimental investigation employing both physical and perceptual measurements was undertaken. Four sophisticated multichannel microphone arrays were optimally placed for simultaneous capture of a series of four piano performances, so that resulting auditory imagery could be compared between otherwise identical performances. In an initial study, preference choices of two groups of listeners showed that not only the microphone techniques in use, but also differences in the musical content, contributed to the modulation of the listeners' preference. In order to predict the obtained preference of the listeners through the analysis of the binaural signals, differing versions of the musical programs were captured through the Head and Torso Simulator (HATS). The analysis showed that a combination of two electroacoustical measures, Ear Signal Incoherence (ESI) and Side Bass Ratio (SBR), was closely related to the obtained preference. The subsequent analysis revealed that the salient perceptual attributes were described by five bipolar pairs of adjectives, and the listener preference was mostly accounted by those attributes. However, the ESI and SBR were not associated with any of the salient attributes, which led another investigation to find the physical measures associated with the perceptual attributes. New physical measures were then derived from the analysis of binaurally captured multichannel reproduced piano music for WIDTH,

BASS TIGHTNESS and SHARPNESS, respectively. The quantitative relation between physical, perceptual and preferential measures derived from this study could be used to predict the preference for other multichannel reproduced piano music.

ABRÉGÉ

Afin de comprendre les caractéristiques latentes de l'enregistrement multicanaux de la musique reproduite d'un piano solo, quatre techniques sophistiquées de captation micro multicanaux ont été déployées de façon optimale pour saisir les mêmes performances simultanément. Les résultats des échells de préférences chez les auditeurs en relation aux quatres techniques de captation multicanaux utilisées démontrent que non seulement les techniques de captation, mais aussi le contenu musical contribuent à faire moduler les préférences des auditeurs. Afin de prédire les préférences obtenues des auditeurs par l'analyse des signaux binauraux, différentes versions des programmes musicaux furent capturés é l'aide dun Simulateur Tête et Torse (HATS). L'analyse a démontré que la combinaison de deux mesures électroacoustiques, le *Ear Signal Incoherence* (ESI) et le *Side Base Ratio* (SBR), était étroitement reliée aux préférences obtenues. Une analyse subséquente a révélé que les principaux attributs de perception ont été décrits par cinq paires d'adjectifs bipolaires et la préférence des auditeurs était principalement représentée par ces attributs. Cependant, l'ESI et le SBR n'étaient associés à aucun des principaux attributs, ce qui conduisit à une nouvelle investigation afin de trouver les mesures physiques associées à la perception des attributs. De nouvelles mesures physiques furent alors dérivées de l'analyse des captations binaurales multicanaux de la musique reproduite d'un piano, soit pour la largeur, la précision des basses et la netteté respectivement. La relation quantitative entre les mesures physiques, perceptuelles et préférentielles dérivées de cette étude

pourrait être utilisée pour prédire la préférence d'autres enregistrements multicanaux de la musique reproduite d'un piano solo.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
Introduction

1.1 Background to the Research

Understanding the complex characteristics of a multichannel sound is a major challenge for a recording engineer or Tonmeister, who wants to create a satisfying sound quality and deliver it to listeners. This topic ranges from designing a recording technique that captures a sound event to delivering the sound via multiple loudspeakers, covering deep knowledge of acoustics, electro-acoustics, and psycho-acoustics.

Recording and delivering a musical performance with sufficient sound quality has been a great challenge ever since Thomas Edison manufactured the first working phonograph.[1] It has been mainly technical limitations that have prevented equipments from capturing the complicated acoustical characteristics, extensive dynamic range variation and wide spatial distribution of musical instruments. Today, with technical improvements to devices that include a wider frequency response and less distortion noise, it has become possible to reproduce the extensive dynamic range of classical instruments such as the piano. However, the authentic impression of a real instrument is quite different from a sound field reproduced through a single channel.

[1] Charles Cros, a French scientist, published a theory of phonograph before Edison.

The first two-channel sound reproduction was evocative enough to give listeners the feeling of being in another space such as a concert hall, and the system came to be called as "stereophonic[2] ." The advancement rested on the fact that the two-channel system made it possible to deliver spatial information that had been limited in single-channel reproduction. While in popular music, this new reproduction system motivated musicians to compose many creative sound fields, in classical music, recording technologies have evolved to create a more authentic impression of the original performance. Many recording engineers tried to find a method to capture the best balanced sound field for the two-channel system, and proposed the use of an array composed of two or three microphones. These arrays, often referred to as stereo microphone techniques, typically created a virtual sound imagery, also known as a phantom source, between the two front speakers by controlling the inter-channel intensity difference, the inter-channel time difference, or both.

With the advent of the DVD-video format, the audio industry saw a new potential to utilize more than two audio channels for enhanced spatial rendering in personal application and launched ITU-R BS.775-1[4], a new standard for multichannel sound recording and reproduction. The most distinct difference with the multichannel sound was that a sound field could "surround" a listener so as to transparently deliver an authentic impression of the sound field to the listener's room. Many recording engineers and researchers introduced various microphone techniques, which are

[2] This terminology was first coined by the research laboratory of Western Electronic [3].

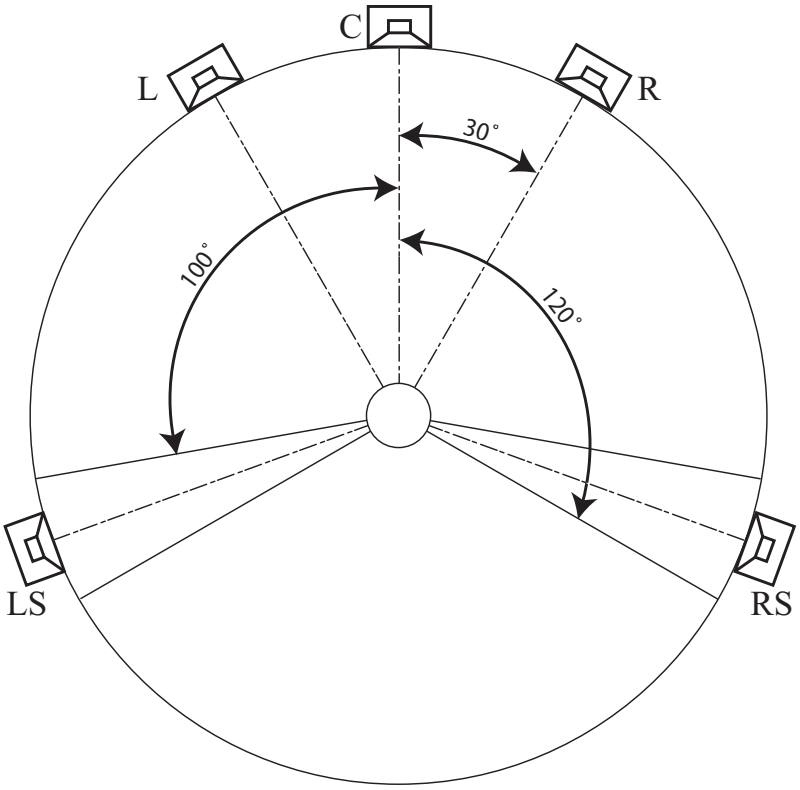


Figure 1–1: Reference loudspeaker arrangement of ITU-R BS.775-1

sometimes called "surround" microphone techniques, that are intended to optimally capture and deliver acoustical events via five speakers, as shown in Figure 1–1. Each microphone technique was based on both theoretical validation in electro-acoustics and the practical appropriateness needed to achieve the required sound quality, while retaining characteristics distinct from its competitors.

The fact that multichannel microphone techniques produce distinctively different characteristics leads the user to consider the use of a microphone technique according to the given acoustical conditions of a venue and instrument(s). Moreover, it is not rare for the recording engineer to optimally adjust the suggested configuration of a

microphone technique to create a sound field that best fits the context. In such a case, the recording engineer or Tonmeister judges the necessity of optimization based on his or her internal reference, which has been formed throughout numerous experiences. It would be valuable to find out what characteristics of the multichannel reproduced sound did influence the engineer's internal reference. Such knowledge would possibly allow the objective controllability required to create and manipulate a sound field that many listeners will favor. To date, few direct comparisons of multichannel microphone techniques compliant with ITU standards have been conducted. In part, this is because multichannel audio is relatively new, but it is also because of the practical difficulty of obtaining the large numbers of high-quality microphones necessary for such comparisons. This study, therefore, compares multichannel microphone arrays compliant with ITU standards that have been optimally positioned for each musical selection in order to maximize the sound quality.

It would be ideal to conduct such a comparative analysis for many instruments, but this is practically impossible. Thus, it is necessary to choose an instrument which can represent a variety of both musical and acoustical characteristics. Among the many instruments, the piano is the one that produces the broadest frequency range, a huge variation in dynamics and the widest radiation pattern. Also, this instrument is known as the most difficult one to capture performance details and reproduce musically [5]. The complexity and variety of physical characteristics make the piano an ideal candidate for this study because the latent perceptual attributes associated with multichannel audio might not be stimulated for an instrument with limited physical variation. For example, a flute sound would not be enough to drive

the spatial variation uniquely identified in the multichannel reproduced sound field. By understanding the perceptual characteristics of the reproduced sound fields of solo piano music, it would be possible to link and apply the results to other instruments. It is worthwhile noting that an ensemble of instruments involves more complex issues such as timbre integration and scene segregation which makes it hard to conduct a systematic evaluation of the multichannel reproduced sound field. The solo piano, therefore, would be the proper sound source to analytically investigate the perceptual attributes and associated sound quality of the multichannel reproduced sound field. However, a reader should caveat that the result from this study has been driven for multichannel reproduction of solo piano music; applying the quantitative equation directly to the surround field of other instrument might require a proper adjustment considering the distinct characters of the instrument.

1.2 Purpose of the Dissertation

The research presented in this dissertation aims to measure both the physical and sensory characteristics of a series of multichannel recordings of solo piano concert music from a variety of stylistic periods, and attempts to relate these measured characteristics to listener preferences for a variety of recordings of a single instrument. Prior related research has focused primarily upon listener preferences for various multichannel microphone techniques, without attempting to discover why one technique produces a result that is preferred to another.

A subsequent goal of this dissertation research project is to relate the sensory characteristics of these same multichannel piano recordings to observed listener preferences.

Finally, the project aims to develop prediction equations that relate physical measures to the auditory attribute ratings associated with the presented multichannel piano reproductions.

This dissertation also scrutinizes the interaction between the three domains mentioned above and aims to build quantitative relationships, hoping that these relationships can be used to analyze and understand the characteristics of any multichannel sound field.

1.3 Structure of the Dissertation

This dissertation consists of ten chapters: Introduction, Literature Review I, Literature Review II, Summary of Related Manuscripts, Three of Published Manuscripts, Contextual Effects, Conclusions, and Contribution of authors in each manuscript in order to deliver the questions for research purposess seamlessly.

The introductory chapter contains the background to this dissertation research, with a short summary of the history of music recording and reproduction.

The following two chapters present an overview of two sub-themes related to the research, which are essential to understanding the main contents of the dissertation; the first reviews and summarizes the theoretical backgrounds of various Multichannel Microphone Techniques for the 5.1 reproduction system; the second reviews the literature related to Sound Quality Evaluation in the context of reproduced sound.

The next chapter contains the summaries of three publications which embody the main contents of this dissertation. This summary chapter explains the core ideas of each experiment, methods used, and results in order to envisage the progress of the entire research and help readers to understand the following manuscripts easily.

The chapter is followed by is a collection of the three main publications listed below.

- An Examination of the Influence of Musical Selection on Listener Preferences for Multichannel Microphone Technique, *Sungyoung Kim, Martha DeFrancisco, Kent Walker, Atsushi Marui, and William L. Martens*, in the proceedings of the 28^{th} International AES conference, June 2006, Piteå, Sweden.
- Predicting Listener Preferences for Surround Microphone Technique through Binaural Analysis of Loudspeaker-Reproduced Piano Performances, *Sungyoung Kim, William L. Martens, Atsushi Marui, and Kent Walker*, in the proceedings of the 121^{st} International AES convention, October 2006, San Francisco, USA.
- Deriving Physical Predictors for Auditory Attribute Ratings Made in Response to Multichannel Music Reproductions, *Sungyoung Kim and William L. Martens*, in the proceedings of the 123^{rd} International AES convention, October 2007, New York, USA.

The following chapter features a discussion on the contextual effect observed in the preference response and perceptual measurement throughout this research.

The next chapter consists of conclusions, a discussion of further related investigations, and future work.

And the final chapter of this dissertation presents contributions of authors of the three main publications.

CHAPTER 2
Overview I: Multichannel Recording Techniques in the five-channel reproduction system

2.1 Introduction

In this chapter, the design criteria and psychoacoustical characteristics of several well-known microphone techniques are reviewed in order to build a foundation on which the following series of experiments in this dissertation can be understood. As previously stated, a multichannel microphone technique refers to an array of multiple microphones configured for optimal capture of an acoustical event, utilizing interchannel intensity difference, interchannel time difference, or both. The two most distinct advantages of multichannel (5.1 or 5.0 channel) audio are the superior localizability and the enhanced spatial impressions such as spaciousness, immersion, and envelopment due to an additional center channel and two rear channels. Subsequently, any multichannel microphone technique is challenged to maximize such benefits of multichannel audio. These advantages, however, could deteriorate a multichannel auditory imagery when inappropriately manipulated. For the systematic evaluation of a multichannel microphone technique, the whole discussion is here limited to the application of the acoustical or classical recording where there can be as few as one microphone signal per delivery channel/loudspeaker and there is a minimal application of processing or added effects after the initial capture of the instruments in the performance space. In popular music or computer music, a microphone array

is a creative and active tool to correlate with musical need. Once again, it would be worthwhile to note that "multichannel" in this dissertation implies the 5 channel reproduction as specified in the ITU-R BS.775 (as shown Figure 1–1) and also that these microphone techniques are most appropriate for use "when the spatial acoustics of the environment are as important as those of the source within[6]", such as in classical music.

2.2 Main microphone array

Multichannel microphone techniques generally fall into two groups - "main microphone array" and "front/rear separation." This distinction is based on whether a microphone array is trying to reconstruct an auditory scene precisely and seamlessly around the listener (360°) or capture the spatial information more effectively, which also has been an important criterion in the two channel stereo microphone techniques[1] . In two-channel stereo microphone techniques, near coincident pairs are known as having balanced the time-intensity combination to produce a relatively accurate sound image with spaciousness, which is based on well-known Williams Curves[7]. Consequently, this compromise between accuracy and spaciousness also became one of the main issues in the design of a multichannel microphone array. An array that tends to achieve optimal time-and-intensity difference among all microphones in use is generally regarded as a "main microphone array." In other words,

[1] In two-channel stereo microphone techniques, precise localization and spaciousness have often conflicted. The two camps are usually represented by either **Coincident** pair such as Blumlein technique or **Spaced** pair such as AB technique respectively.

a multichannel main microphone array tries to make five *critically linked* sectors decided by each pair of two adjacent microphones. Before going further it would be beneficial to understand this Critical Link concretely. This concept has been proposed and elaborated upon by William Michaels and basically is based on the idea that a pair of two microphones create an auditory imagery between two associated loudspeakers, and the horizontal extent and distribution will be decided by the distance, the angle, and the directivity of the two microphones. Once a pair is set to have optimal extent and distribution of the associated spatial imagery, it can connect to its adjacent (also optimized) imagery. When these two adjacent imageries are formed a continuous "link" so as to distribute all auditory components without spatial distortion, it can be said that the two sectors are *critically linked*. Williams and Le Dû proposed various configuration to obtain this critically linked microphone array using cardioid. INA3 and INA5 are two good commercially available multichannel microphone arrays in this group. INA stands for Ideale Nieren-Anordnung (this can be translated to Ideal Cardioid Arrangement) and was developed by Hermann and Henkels. One important drawback of this approach is that the pleasing listening area (or so-called sweet-spot) is relatively narrow because optimal localization is built for a listener who is in the center of the listening position which is equidistant to all loudspeakers.

The other similar but distinct approach is to use a SoundField microphone based on Ambisonics and extract a multichannel feed, also known as G-format, from its original output B-format [8]. Ambisonics is a sound recording and reproduction system whereby a sound field which exists at a point in the recording space is recreated

at a point in the listening environment. A SoundField microphone comprises four sub-cardioid capsules and these sub-cardioid signals then transform to B-format signals via a dedicated convertor which comes with the microphone unit. B-format signals are equivalent to four coincident microphones: three figure-of-eight mics and one omni-directional mic which correspond to X, Y, Z and W. W is called zeroth order, X, Y, and Z are first order, which is also known as B-format. The four signals are then decoded to any required loudspeaker configuration, located at arbitrary locations in the listening room. It is possible to have full periphonic (with height reproduction) with this system. There is a dedicated B-format to 5.1 channel decoder (as explained in [9]), which is broadly used for the sound effect of movie and broadcasting. However, compared to other multichannel recording methods, its reproduced imageries were reported as narrow and distant [10], which is most likely due to the strong correlation between channels.

2.3 Separating front and rear segments

The current standard of multichannel reproduction format does not intend to enable perfect localization of horizontal space around a listener. If that was the goal, more loudspeakers should have been symmetrically placed. This has inherited the custom of multichannel audio in theatrical use - the rear or surround channels have been employed to deliver the sound effects which helped listeners to immerse themselves in the scene. This approach can be applied for classical music reproduction as well. When we refer to classical music, it generally implies a staged music where instruments play in front of a listener and environmental sounds come from the side and rear. Many multichannel microphone techniques for classical music and natural

sound, therefore, have been proposed to capture front sound stage and rear sound stage separately but more efficiently.

2.3.1 Frontal array

The design criteria for the front three microphones is generally the same as for various microphone techniques, which is to obtain precise localization and better depth perception of the frontal sound field. To obtain such effects, the frontal three microphones are placed to achieve a "Critical link" in the reproduction stage. Fukada from NHK proposed to adopt three identical cardioid microphones and two omni-directional microphones as a frontal array [11]. The functionality of these two additional "omni" mics will be explained below with regard to the integration with the rear array. Klepko[12] from McGill University proposed to adopt two super-cardioids and one cardioid to prevent a "skewed centre buildup" due to the overlap between phantom center from the left/right microphones and real center from the center microphone. Theile [13] investigated this issue in depth and claimed that the interchannel crosstalk should be minimized. He subsequently proposed a microphone array that incorporates two hyper-cardioid microphones for left and right channel in addition to one cardioid microphone for the center, which is called OCT (Optimized Cardioid Triangle).

An effect of interchannel crosstalk (ICC) fed from an adjacent microphone is one of the debated issues in the use of a multichannel microphone technique. Proponents who give a strong focus on the artifact of this ICC asserted that a reasonably balanced localization could be possible when the intensity of ICC was as reduced as possible. According to their hypothesis, with a relatively large ICC, it would be possible

to have triple phantom images corresponding to the three front channels due to the similarity between signals. However, this hypothesis of triple phantom images has not yet been supported by any experimental evidence and is also debated by opponents who state that a listener tends to perceive a single fused phantom image dependent on the relative intensity and time differences between the signals rather than perceive triple phantom images. The use of a spaced microphone array also reduces the consequence of crosstalk due to the precedence effect. In later investigation [14], it has been shown that the most perceptually salient effects due to interchannel cross talk have been reported as a function of an increase of *source width* and decrease of *locatedness* decided by the ratio of time and intensity differences in a microphone array.

In a recent investigation into multichannel microphone technique, Martin [15] asserted for use of *coherence* instead of a correlation coefficient as a metric to measure the similarity between two audio signals. This coherence between two signals can be regarded as a set of correlation coefficients at different frequencies. In this paper the author showed that relatively high interchannel coherence is required to obtain wide variation in IAID (InterAural Intensity Difference) in the front sound field, which is counterintuitive against the general idea of a relationship between interchannel coherence and interaural difference. The model used his paper to show the relationship, however, has one big assumption (as the author admitted): constant level difference to constant directivity of a microphone over frequency. Therefore it would be interesting to measure whether this experimental and model-based result would coincide with the in-situ measurement.

In general, in the design of a frontal array for the multichannel microphone technique, it would be wise to approach it based on the concept of "critical link" to obtain better localization through the frontal array. A solid and continuous frontal array might be built with significant interchannel crosstalk and corresponding high correlations among channels.

2.3.2 Rear Array

As previously mentioned, when a listener hears a stage music, most environmental sound comes from the side and rear while instruments onstage come from the front. This idea extends that the rear microphones of a multichannel microphone technique should be designed to capture the environment enough to recreate the immersive impression of "being there." It is usual for a concert hall to have a bigger volume than a listening room. Therefore, it is necessary to create the increased spaciousness in order to recreate the original environment. In general, such feeling of spaciousness or diffuseness can be increased when the signals fed to loudspeakers are decorrelated. However, when decorrelation between two speakers is over certain degree, reproduced sound field might be discontinued and be localized towards the position of the loudspeakers themselves.

The developer of a microphone array now have to deal with the paradox that decorrelations between channels should increase but at the same time should not go over a certain degree. In other words, it is therefore required for the designer of a microphone array to obtain a rear microphone array the results of which are decorrelated enough to create diffuseness but at the same time not enough to lose

the smooth transition from the frontal array to the rear array and the same smooth transition between the two rear channels.

Most multichannel microphone techniques consider this aspect. In the Fukada tree, two cardioid microphones were used to reduce the direct sound component by facing those microphones backwards to audiences (not onstage musicians or instruments). And the angle and the space between the two rear microphones were large enough to decorrelate the two rear signals. Since this rear array is intended to capture the environmental sounds, it is located beyond the critical distance of the hall or space. In order to give connection between the front and rear arrays, which also allows the sound stage to envelop the listener, the Fukada tree adopts two omni-directional microphones which are placed outside of front left and front right microphones. These two microphones not only help to enlarge the coverage angle for wide instrumentation such as an orchestra but also to connect the front and rear arrays by delivering lateral energy in a concert hall.

The two most frequently used rear arrays are IRT-Cross and Hamasaki Square. These two arrays share the same design criteria: four microphones capture the four decorrelated components associated with a recording venue and reproduce them through four loudspeakers - left, right, left-surround and right-surround speakers. IRT-Cross adopts four cardioid microphones quadratically distributed at ± 45 and ± 135 from the center front of a listener. Hamasaki Square [16] adopts four bi-directional microphones facing its positive direction of diaphragms toward the walls at each side. Hamasaki Square suppresses the direct sound component much more

and increases the lateral energy efficiently, which allows the required diffuseness and pleasing envelopment simultaneously.

Klepko [12] proposed to use HATS (Head And Torso Simulator) and corresponding binaural signals for the rear channels. Since two rear loudspeakers are located perpendicular to a listener's ear, it would be possible to deliver binaural signals with a negligible amount of crosstalk to the contralateral ear. After equalization is applied to get rid of the double pinna effect, these binaural signals can reconstruct the environmental sounds around the listener efficiently. However, this technique generates a very narrow listening area.

As a summary, it can be said that a rear array should be designed so as to deliver the effect of the rear channels - spaciousness, presence, diffuseness, etc. - without actually hearing their presence [17].

2.3.3 Downward compatibility

Gernemann [18] proposed a frontal array which considers the compatibility to two-channel stereo reproduction. His method adopts a stereo pair for left and right speakers and lets this pair to handle sound stage, and with an additional microphone which is located sufficient distance from the main pair so that the precedence effect is observed and only a phantom source can be generated. By doing so, it is possible to take only the left and right channels when needed to reproduce in a two-channel playback environment. However, this approach conflicts with other methods which try to utilize the benefits of multichannel audio as Wuttke pointed out [17] as below,

> *... Thus the center channel should play such an important role that it becomes truly indispensable, even in audio-only applications...*

It does not mean that a designer of a multichannel microphone array should not consider the downward compatibility, rather it points out that a multichannel microphone array should have the distinct ability to enhance the experience through multichannel reproduction and let the compatibility issue be handled by other professionals either by applying proper downmix or by producing a separate stereo record.

CHAPTER 3
Overview II: Sound Quality Evaluation

3.1 Introduction

Much research related to sound and auditory information has delved into understanding the underlying structure of human perception, associated with its perceived quality. The sound quality evaluation technique has been developed as a primary tool for that purpose. In the literature of acoustics, this technique has been served as a subjective assessment of auditory information from a performed sound field in a venue [19, 20, 21, 22]. Recently it has been applied for the evaluation of the reproduced sound field created and delivered by electro-acoustical equipment. In particular, the quality of speech has been significantly investigated with the remarkable growth of the telecommunication and mobile communication [23]. As the new paradigm of multichannel audio has allowed a distinct sound field from its predecessor, two-channel reproduction, many recent research projects have applied quality evaluation methods to understand the distinct characters and quality of the new multichannel sound field [24, 25, 26, 27]. This chapter will introduce and summarize the current sound quality evaluation methodology in order to better understand the research methods adapted in this dissertation research.

3.2 Definitions of sound quality

Sound quality is a multi-faceted attribute defined by various physical and psychological aspects. In the early era of sound and audio technology, the quality of

sound was often regarded as equivalent to the quality of a device, which referred to the collection of its technical descriptions such as signal-to-noise ratio (S/N) and magnitude response. As Soren and Zacharov stated in their book, while such a parametric profile could explain associated characters of the device, it "does not tell us how the human auditory system will interpret and quantify it [28]." This is because the final receiver of any sound (either performed or reproduced) is a human auditory system [29]. Therefore, even though the profile of technical descriptions could deliver significant information about the sound field, it would be more important to understand the response of a listener to the exposed sound field. Considering the importance of subjective response, Letowski defined sound quality as a result of "(global) assessment of auditory image in terms of which the listener can express satisfaction or dissatisfaction with that image" [30]. He also asserted that sound quality is a different concept than a sound character which is supposed to be purely descriptive. Letowski then viewed sound quality as an integration of sound characters inherently unique in the sound field. Later, Blauert quoted Jekosch's definition of sound quality of a speech in [31] :

> *... Speech quality is the result of an assessment of the adequacy of a speech sample - considering all of its recognized and nameable features and feature values namely, as to which amount this speech sample complies with a reference arising from aspects such as individual expectations and/or social demands and/or pragmatic necessaries - considering all recognized and nameable features and feature values of the reference...*

It was the degree of an adequacy to the reference that Jekosch and Blauert regarded as the magnitude of quality. The difference between the two definitions (Letowski vs. Blauert) showed that sound quality can be differently interpreted depending on the given application, which subsequently makes it hard to replace sound quality with a certain percept such as satisfaction or adequacy. Considering this fact, Rumsey defined sound quality as a "composite entity ... that conflates all aspects of sound quality, including preferences and descriptive characteristics, into a single rating [32]." What is shared by these three definitions is that sound quality affects listeners' response; it affects so that the listeners choose a specific sound field against its competitors, and judge relative adequacy for a certain application. Therefore, sound quality can be defined as an affective response that is an integration of percepts created by sound reaching at the listeners' ears, and that influences their sentimental, preferential, and judgmental reactions.

3.3 Methods

The purpose of sound quality evaluation is not limited by the observation of the listeners' affective response to a sound field; rather it endeavors to disclose the independent factors that influence that affective response. It has been generally accepted that the overall affective response can be decomposed to the descriptions of sound characters and the associated perceptual measurements. Further, the measured quantity of the percepts, including the affective response, can be mapped to the related physical measurements.

Previous investigations in sensory evaluation have separated the overall affective response into the three sub-domains as shown in Figure 3–1. This concept has

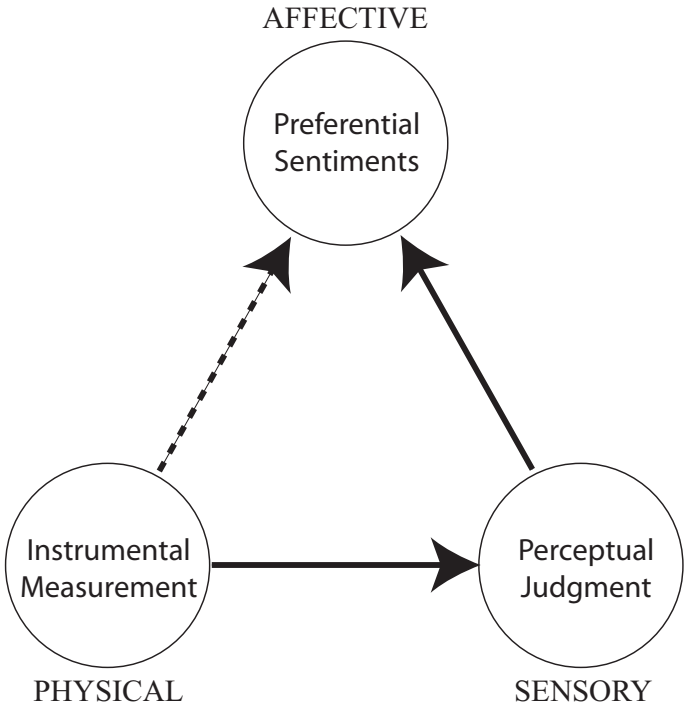


Figure 3–1: Three components that play a role in the general assessment of perceptual evaluation

been widely adapted among many sensory evaluation studies such as food and scent sciences [33, 34]. Hence, a sound quality evaluation refers to a scientific research method to systematically analyze the underlying relation among physical, sensory, and affective domains in an auditory stimulus, and derive a model that would result in a quantitative estimation of affective response.

Bech [1] expanded and specified the relation of the three domains and created a more detailed diagram for sound quality evaluation shown in Figure 3–2. He also classified the diagram into three major steps required for the process of the evaluation of a sound field:

1. To identify the individual auditory attributes
2. To devise methods for obtaining a measure of the magnitude of sensation for each attribute
3. To establish the relation between the auditory attributes and the total impression

This is a commonly adapted method in a current sound quality evaluation. Later, Martin and Bech revised this approach and suggested a two-filter model for the automobile sound quality evaluation [35]. Nonetheless, it is not rare to find the literature to eliminate the sensory domain and make a direct connection between the physical measure and sound quality [36, 37] which is the dashed arrow shown in Figure 3–1. Such an approach is still valuable for the engineering control so as to contain the required physical characters for a product. However, it is hard to explain why those physical characters or measures have affected the total impression. Stone and Sidel wrote " ... optimizing product preference (has) only been possible with

use of descriptive analysis to identify the specific sensory characteristics ..." in their book [34] and asserted the importance of the sensory domain in any quality-related study. Thus a complete model of a sound quality evaluation needs to follow the three steps that Bech proposed as done in other studies [27, 35, 38].

3.4 Attribute Identification

A fundamental assumption underlying modern techniques for sound quality evaluation is that sound quality is composed of sound characters or separable attributes. At the same time, there is a fundamental difficulty which is the problem of identifying the salient attributes of a set of sound stimuli, and defining those attributes using language that an untrained listener might readily understand. There have been two major classifications regarding elicitation of attributes in the sound field: verbal and nonverbal methods.

Verbal elicitation methods have been used in the perceptual evaluation of reproduced sound on the grounds of an assumption that when an auditory stimulus is perceived, the associated character of that stimulus can be identified and related to a set of verbal descriptors. Perceived character of a stimulus is compared to a list of words in memory in order to find a descriptor that might be similar to the subjects' current experience. As Mason *et al.* wrote, "with language we make sense of auditory events (stimuli), translating these events into a meaningful set of terms in order that we may communicate effectively what we have heard [39]." Verbal elicitation methods are powerful and effective tools in formal and informal sensory evaluation because people use language as a main device to communicate and share perceived information [34, Chap. 6][25].

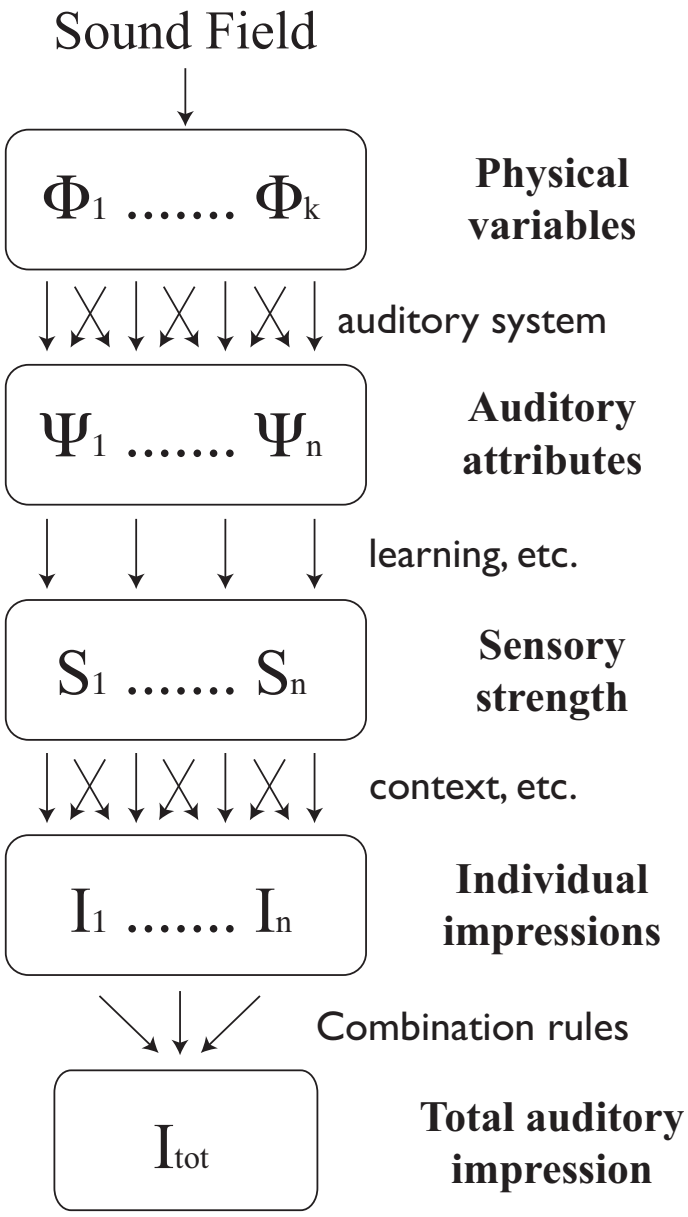


Figure 3–2: The processes of a sound quality evaluation, recreated from [1] with permission of the author

Verbal methods have several artifacts as Mason *et al.* summarized in their paper [39]. First of all, anomalies in verbal communication occur as a result of the symbolism of language (categorizing and symbolizing the object with the word, instead of describing the object), the knowledge and personal histories of the communicators, and the context of the communication. These anomalies can cause the language in use to have multidimensionality or distortion in semantic space. Secondly, there are not enough words available to express all experience. And finally, interpretation at the receiver-side can also be strongly context-dependent. Therefore, it can be said that a verbal description is only effective when there exists common understanding between a communicator and a receiver. As Letowski surmises, the sheer number of terms used when describing a sound "is a blessing for artistic freedom, but a problem when it comes to meaningful communication between people [30]."

To reduce the drawbacks of verbal communications mentioned above, several sophisticated methods have been proposed and modified. Among them, one of the most commonly used in the current verbal elicitation methods is Descriptive Analysis (DA). DA is "a sensory methodology that provides quantitative descriptions of products, based on perceptions of (a group of) qualified subjects [34]." Qualification of subjects implies that people who take part in the elicitation experience have a certain period of training either as a group or individually to express and share perceived attributes via common language. DA, therefore, essentially requires the majority of subjects to generate and agree upon a common set of languages for the given stimuli. In this process, two important things should be considered in order to avoid the ambiguity of elicited verbal descriptors as Martin and Bech pointed out

[35]. First, emotional or attitudinal languages should not be used since these are too context-dependent and hardly objectively quantified. Second, chosen descriptors should have uni-dimensionality in semantic space. In other words, elicited descriptors should not covary and should remain as orthogonal to each other as possible. In many cases, descriptors composed of two bipolar adjectives with generally opponent concepts (such as antonyms) are used. Estimation of the perceived magnitude associated with each attribute allows the experimenter to find interrelationships among elicited attributes and moreover to connect these attributes to overall quality.

Non-verbal methods are, as the name implies, an identification tool not based on the linguistic response of the brain, but rather based on the motor-visual response such as drawing and pointing. The general assumption of these methods is that within the cognitive mechanism, there is one system that allows perceived characters to be recognized with languages and another to be recognized with mental imagery. These two systems and their functionality are together known as a dual-coding system [40]. Two internal systems separately accept the stimulus and stimulate the associated verbal semantic and mental imagery but at the same time these are interconnected. In other words, the two systems are functionally independent but interconnected so as to support each other when one system fails to match the description to the stimuli. There are also several characters of stimuli that can be processed through both systems, but other characters are likely to be much more effective in one system than another, resulting in fewer cognitive loadings and chances for misinterpretation. Therefore, non-verbal methods can be effectively used for the several perceptual attributes for which verbal descriptions are either not available or

too multidimensional to communicate and interpret. In particular spatial attributes in auditory evaluation can be represented and quantified effectively and with less variance by using non-verbal methods [41, 42, 43, 44, 45].

However, non-verbal elicitation methods also have drawbacks that require them to be used with care. A common problem is dealing with the subjects' perspective. This problem stems from the fact that many sensory evaluation experiments require the subject to project three-dimensional phenomena into a two-dimensional plan. One listener can have imagery that is seen from the bottom while another can have a view from the top. In other words, it is relatively hard to standardize external reference points where, if not specified, a subject usually takes his egocentricity [39]. Even less than verbal methods, non-verbal methods are also dependent on familiarity with the context. It is clear that these non-verbal methods are selective in the information they can provide, and the results are open to interpretation. Therefore, if the experimenter wants to extensively examine the structure of auditory perception, it is advisable that both verbal and nonverbal methods be used, especially in the evaluation of reproduced sound.

3.5 Perceptual measurements

Whereas identifying the salient perceptual attributes reveals the inherent characteristics of the given sound field, the interrelation underlying the inherent characteristics and the affective response is often analyzed through the measured quantity; Stone and Sidel asserted that [34, Chap. 3] measurement is "*critical to quantifying responses to stimuli for the purpose of utilizing descriptive and inferential statistics.*" Measuring sensation has been profoundly investigated by early researchers including

Weber, Fechner, and Stevens to whom the modern sensory studies are in debt (see [46] for the summary of the early works by Stevens). A good example of the importance of perceptual measurement can be found in a randomly chosen issue of a journal that contained "all 12 major articles in the issue reporting on studies that involved the use of scales" (quoted from [47]). Consequently, a meaningful measure of a subject's response has been a major task in many disciplines including psychology, psychophysics, and psychoacoustics. Modern psychoacoustical research has focused on the observation of the **facts** (as shown in the title of Zwicker and Fastl's book [48]) between the perceptual and physical domains by comparing the measured quantities of the two. An experimenter should know that, unlike a physical measurement, any perceptual measurement is relatively time-consuming and more likely to be inconsistent. Therefore, it is important to determine a proper experimental method of measurement by which the listener's response could convert into a meaningful metric. There are two main categories in the measurement of perceptual responses: (1) indirect methods and (2) direct methods.

3.5.1 Indirect methods

An indirect method test does not ask the listeners to map the perceived magnitude of a percept "directly" to the associated numeric value; rather this method requires the listeners to detect the percept in a test and forms a scale from the proportion of the detection based on the statistical analysis. The proportion of the detection will form two types of psychoacoustical measurements: (1) Just Noticeable Difference (JND) and (2) Continuum of the perceptual magnitude.

The detecting threshold of a percept and building a JND has been a major topic in psychophysical and psychoacoustical studies [49, 48] as well as audio engineering research [50, 51, 52]. While it is possible to detect a JND using the method of adjustment, indirect methods are often used to find a detection threshold via two ways: the constant stimuli and the constant response [49]. One interesting aspect is that it is possible to build a continuous scale from the series of JND. Bech and Zacharov wrote that scales are "determined experimentally as a function of stimulus intensity by successively increasing the stimulus intensity by one Just Noticeable Difference (JND) [28, Chap. 4.2]." While the method based on JND detection has served as a major tool of psychoacoustical studies, one drawback of the method is the preparation of the variation of the stimulus' intensity, which might not be possible for in certain experimental conditions.

In contrast to the method based on JND detection, the method based on comparison allows building a continuous scale of a perceptual magnitude of the stimuli from the frequency of selection. The listeners compare multiple stimuli and choose one which has the most or the least magnitude of the attribute in test. This method requires a solution to the problem of expressing experimentally observed dominance proportions as a function of underlying scale values for the stimuli. Among many methods proposed by researchers in many disciplines (see [53] for the related bibliography), Thurstone [54] proposed various models of converting choices to the associated scale with the particular assumption in the experiment. In his terminologies, a pairwise choice is a result of the "*process by which we react differently*" to the given stimuli and is called "*the discriminal process.*" The simplest form of Thurstone's law

of comparative judgments (Case V) assumes equal variances for the discriminal processes. With this assumption satisfied, Case V of Thurstone's law can produce the "*psychological continuum*" of a perceptual attribute. A substantial amount of recent references showed that their perceptual measurements were scaled by Thurstone's law from pairwise choices [55] [56] [57] [58] [59].

A modern approach to convert the choices to the scale is based on the Bradley, Terry, and Luce (BTL) model [60, 61]. Several recent studies [62, 63, 26] adapted this method to create a scale related with the listeners' hedonic response. Wickelmaier and Schmid [64] have provided a Matlab function that constructs scale values from pairwise preference choice data, which has been adapted in this dissertation research. Martens *et al.* summarized the comparative characteristics of the law of comparative judgment and the BTL model:

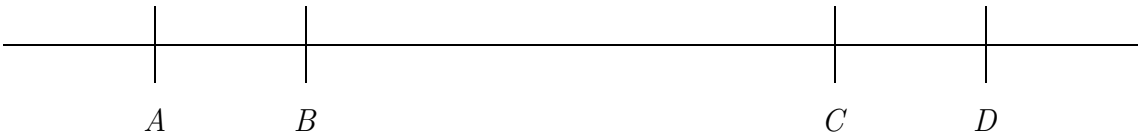
> *Whereas Thurstone regarded variability in response to be due to variability in perception for each stimulus, Bradley and Terry's model attribute it to variability in listener judgments. Also, Thurstone's model is based on the assumption that variability in perception for each stimulus follows a normal distribution, whereas, Bradley and Terry took binary choice data as a special case of ranking multiple stimuli, and used the binomial distribution as their representation for choices of one stimulus over another. Although the approaches are different, the scale values obtained in the two methods are similar in many cases.* [65, Section 2.1.]

While the pairwise comparison is used relatively often to measure the affective response of the listeners, Choisel and Wickelmaier used the indirect method to build a

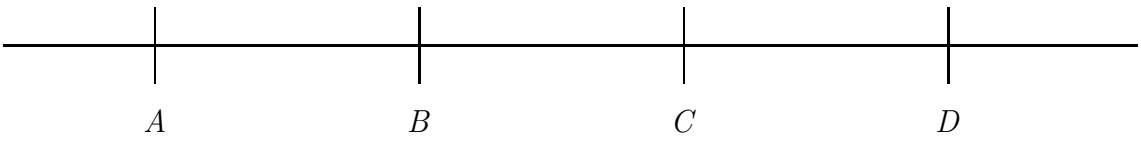
scale for each of the salient attributes in their study [26]. Thomas [66] recommended to conduct the pairwise comparison to measure the overall impression by a large number of untrained listeners and collect the direct measurement of specific attributes by a small number of trained listeners.

3.5.2 Direct methods

Whereas a scale generated through an indirect method produces a highly reliable result, the cost of the experiment is much higher. For example, in order to obtain a reliable scale, more than thirty observations or comparisons are often necessary. Another drawback of an indirect method is that the conversion of the proportion could mislead an erroneous scale when the differences between stimuli are obvious. For example, let's assume that there is a hypothetical scale that we want to extract through an indirect method as below:



If the perceived difference between A and B, or the difference between C and D, is large enough, then the listeners always choose B over A and D over C. Even though the difference between B and C is much larger than the two aforementioned differences, the listeners choose C over B with the same proportion as choose D over C. Therefore the converted scale from the indirect method would look more like the one below:



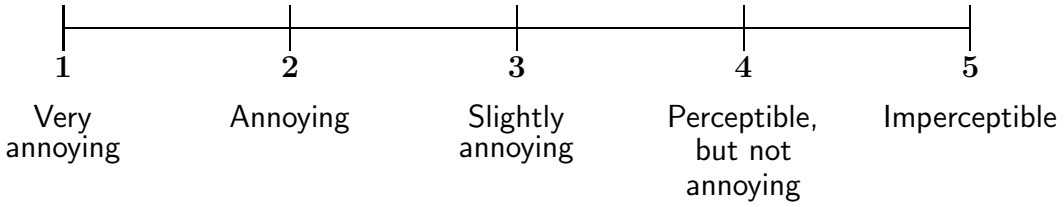
Hence, a direct method might build the required quantification more effectively and precisely if the perceptual difference among stimuli is relatively large and the listeners are aware of the task and the percepts. A direct method is an experimental design to collect the magnitude of a percept as a result of a direct conversion to a metric by a subject. Since it is based purely on the internal mapping processed in a brain, it can be affected by non-experimental variables, which often makes it hard for a normal listener to respond consistently. Regarding the context effect in the rating process, the Chapter 8 will cover the related contents.

Stevens' four level of measurements [67] show the classical and basic classifications of scales associated with the different information. These might serve as the first guideline for the experimenter to choose a proper scale for the given task. Stevens four categories are:

1. Nominal scales for use in classification or naming
2. Ordinal scales for use in ordering or ranking
3. Interval scales for use in measuring magnitudes, assuming equal distances between points on the scale
4. Ratio scales for use in measuring magnitudes, assuming equality of ratios between points

Further, Stone and Sidel [34] summarized and itemized the requirements of a scale in a sensory evaluation: it should be (1) meaningful to subjects, (2) uncomplicated to use, (3) unbiased, (4) relevant, (5) sensitive to differences, and should (6) provide for a variety of statistical analyses. In particular, the result of a statistical analysis is dependent on and limited to the precision of collected data with assumptions of the data distribution. Thus, the experimenter should understand the nature of the task when a direct scaling method is used.

Also, it is of importance to build a clear definition, and anchors if possible, by which any listener can judge the perceptual magnitude and convert it into the corresponding scale with ease. For example, the ITU-R 5-grade impairment scale given in ITU-R BS.1284 [68] requires a listener to both detect the difference and the judge the amount of degradation, increasing the complexity of the task and possibly resulting in unreliable quantification.



Among various types of scales, the measurement based on an assumed-interval scale is pragmatically favored because a quantitative statistical analysis can be applied for it [28]. Once again, while a direct method is relatively straightforward to use, it is yet vulnerable to biases unless it was appropriately applied considering the characteristics of the scale. Since it is beyond the scope of this dissertation to cover all

related issues of identification and measurement of the perceptual attributes, readers should refer to [47], [34], [69], [70], [33], and [28, Chap. 4] for in-depth information and related examples.

3.6 Physical measurement

In many sensory evaluation studies, a physical measurement often refers the overall process to find a physical character that covaries with the empirically measured magnitude of a percept. And it is a quantification process of a percept based on an instrumental measurement that, in turn, can create an arbitrary stimulus that has an equivalent magnitude of a percept. Noble [71] similarly defines this physical or instrumental measure as "in which the property is assessed by a device which imitates the way in which humans perceive the sensory property." The motivation for developing these instrumental measurements is ultimately to replace time-consuming perceptual measurements. However, while several prediction models have predicted the variance of the associated perceptual magnitude with precision, models that have attempted to predict the total impression or affective response have not been successful. Consequently, many sensory evaluation studies have focused on finding the physical measures of the associated perceptual attribute. Many such studies have already well documented how to predict a perceptual magnitude from instrumental measurement(s) (for example, predictors for loudness and sharpness [48][2]). This section introduces how to devise a physical model of an attribute, Auditory Source Width (ASW), as an example. In particular, ASW has been a significant attribute both in concert hall acoustics and multichannel reproduced sound field, which appeared to be one of salient attributes of this dissertation research.

ASW refers to the horizontal extent of perceived sound objects in a concert hall or in a sound field reproduced by a headphone or loudspeakers. The magnitude of this attribute is modulated by the variation of (1) lateral energy fraction (the ratio of lateral to frontal energy), (2) loudness, (3) interaural cross correlation, and (4) frequency region [72]. When a slightly delayed version of the original sound arrives at a listeners' ear with the original sound, the listener perceives it as either a timbral change or spatial modification. And when the delayed signal causes a binaural difference in time, it usually corresponds to the spatial extension of the source (only until the amount of delay is less than around 30ms after which it becomes a discrete echo). While the first reflections caused by walls, doors or ceiling contribute to the spatial extension of the perceived source, lateral reflections contribute more significantly in the variation of ASW, which has been tested by a controlled experiment [73]. Not all of the lateral energy fraction contributes to the modulation of ASW. The lateral energy fraction can be divided into early lateral fraction and late lateral fraction: impulse responses (of a room or a hall) within 80ms from the direct sound (early lateral energy fraction) are relevant to the perceived magnitude of ASW.

The second factor that modulates the perceived ASW is loudness. If one subject listens and compares two identical sound sources that are different in loudness, he is likely to respond that the louder sound is wider. Further, ASW has frequency-dependent characteristics; it varies according to the frequency region in tests. Morimoto and Maekawa [72] showed that ASW was frequency dependent when IACC (InterAural Cross Correlation), loudness, and frequency lateral energy fraction were held constant. Mason *et al.* [74] also conducted similar experiments and concluded

that when IACC is equal to 1, ASW varied with the frequency region in tests (please compare Figure 1. of Morimoto and Maekawa and Figure. 2 of Mason *et al.*). Mid-frequency range (2-4kHz - according to fig 2. of Mason *et al.*) was not producing the same perceived ASW as its flanking bands (lower and higher regions). In other words, equally loud low-frequency signals created a wider impression than mid-frequency. In general, it has been known that high frequency signals have their own characteristic behavior; they did not contribute to the ASW because of the breakdown of phase locking. When this phase unlocking is compensated by the half-wave rectifier and low pass filtering (a sixth-order Butterworth), it then starts to affect the perceived ASW.

More than any other physical parameters which contribute to the deviation of ASW, IACC has been the most related parameter. IACC is a measurement of similarity between two signals that impinge on each eardrum. IACC is usually measured by a dummy head microphone with and without torso simulator or by a small microphone placed in the ear carnal (at both ears) of a human subject. For the concert hall acoustics, IACC has been used as the sole predictor of ASW of the venue; Morimoto and Iida [75] found that the degree of interaural cross correlation (DICC) was correlated with the variation of ASW, which was measured by the KEMAR dummy head without artificial ear simulators using an A-weighting. Later experiments, however, showed that IACC did not only differentiate the ASW; it also modulates other perceptual attributes such as perceived direction, perceived diffuseness, etc. In particular, for wideband complex signals, overall IACC measured on entire frequency

and averaged over entire duration did not correlate with the perceived ASW; rather the perceived ASW has been more closely dependent on the frequency and loudness.

Since ASW perception is frequency-dependent, it has been asked whether it would be possible to manipulate a metric that better predicts the variation of ASW by splitting the signal into small bands, calculating each band's IACC, and manipulating a weighted sum of each band's IACC. The subsequent questions are: (1) how to define weight for each frequency band and (2) how to divide the frequency region. About frequency regions bandwidth, ISO-3382 suggests to use 1-octave band filters. But the recent experiment by Morimoto and Iida showed that 1/3 octave give better predictions than 1 octave band [76]. As many other psychoacoustical models were based on the use of the Critical Bands (CB), it might be beneficial to adopt the CB (or CB-equivalent bands such as the equivalent rectangular band (ERB)). Mason*et al.* [74] calculated the frequency weighting when IACC equals 1, which can be used to estimate the overall ASW. It is still in question whether the weighting function extracted when IACC equals 1 can also be effective for a general purpose. Ueda *et al.* [77] also studied and reported related to the weighting function of frequency but it is limited to the weight factors of two adjacent bands. This showed that weighting factors are also dependent on the band levels. Therefore, an elaborated weighting function should be devised empirically.

Temporal characteristics are also important to consider; research related to finding the duration to measure IACC that can represent ASW of the stimulus and the influence of transient portion are in investigation [78]. One remark is that it has been

known that our auditory system does frequency weighting process calculations before the temporal effect is accounted for. In other words, it can be assumed that the spectral summation comes first and the temporal summation follows [48]. Finding a prediction model of ASW is an integrative work that covers the frequency, temporal, and loudness of the given stimuli.

3.7 Statistical Analysis

Statistical analysis connects the affective, perceptual, and physical, domains in a sound quality evaluation and extracts the quantitative relation among the three. Since many researchers are interested in building a model that can account for the variance of the affective response, sound quality in a broad concept, they applied various multivariate statistical analysis techniques. Multivariate statistical analysis, in general, investigates the relationships between multiple dependent variables and multiple independent variables. In its most usual form, the relationship between a single dependent variable and one or more independent variables is normally sought through experiments. In the sound quality evaluation, the quantified affective response is a dependent variable and others are independent variables. Again, a single multivariate statistical analysis can easily cover the whole content of a book. Therefore, this chapter will only give a very brief introduction of each technique. A reader should refer to the related publications (such as [79] and [80]) when a specific technique is used to analyze the obtained data from an experiment. Based on the collected data type, a multivariate statistical analysis can be divided into parametric and non-parametric methods.

The parametric method aims to predict the character of the population from where data has been collected. Its usual data type is interval and normal distribution is assumed. One good example of the parametric method is multiple regression analysis (MRA). MRA estimates the potential relationship between a dependent variable with multiple independent variables. That estimation usually is shown as the accounted variance, R^2. This statistic refers to the variance that is accounted for by the combination of weighted independent variables. These weights are known as standardized beta coefficients or simply betas, which refer to the individual contribution of each independent variable. In a sound quality evaluation, MRA is often used to reveal how the variances of independent variables, either perceptual or physical attributes, account for the overall variance of the affective response (see [36, 37, 81]).

ANOVA (Analysis of Variance) is a special form of MRA that has categorical independent variables; it investigates the difference in mean between categorical independent variables. When the number of independent variables is N, then it becomes N-way ANOVA. By transforming these categorical variables to dummy variables, an MRA-based analysis can also be made. In many cases, ANOVA generates the F-statistic from which the significance of difference between given categories can be tested. In other words, ANOVA can test whether the listeners' response is significantly different or not, depending on the two or more test conditions. When the relationship between multiple categorical independent variables and multiple quantitative dependent variables is in question, Multivariate ANOVA (MANOVA) can be used to analyze the relationship. For example, the influence of gender and age on

the composite variable of the willingness rating of purchase and the preference rating can be investigated through MANOVA. Whether the multivariate means (means from a combined variable of purchase will and preference) are different or not according to gender, age and the interaction of the two can be analyzed. In contrast, when an experimenter wants to investigate the relationship between multiple categorical dependent variables and multiple quantitative independent variables, then the experiment uses the method called Discriminant Analysis.

The non-parametric method does not require statistical assumptions to be obeyed but gives inherently less statistical power. This method is based on the categorical data and the frequency that a given stimulus was selected over other stimulus. So it investigates the relationship between categorical independent variables and categorical dependent variables. It requires generating the chi-square statistics based on the following equations [79] (when O refers to Observed frequency and E to Expected frequency):

$$\sum \frac{(O-E)^2}{E} \tag{3.1}$$

If this chi-square value is greater than the preliminarily set level, alpha, to prevent type one errors, then it can conclude that the difference between expected frequency and observed frequency is not due to chance.

One purpose for using multivariate statistical analysis is to investigate the relationship between variables and re-configure its equivalent but less dimensional space. Principle Component Analysis (PCA) is the method that investigates the interrelationship between sets of variables and re-orients the position of stimuli based on its

principal components. These principal components have been obtained such that the first Principal Component (PC) can account for the largest amount of variance in the data set and the second PC can account for the second largest amount of variance, and so on. Each PC is orthogonal to other PCs. The PCA results not only re-oriented coordinates of stimuli in PC (SCORES) but also can return the loadings of each variable (COEFFICIENTS). Various audio engineering experiments have adapted and utilized PCA to effectively represent through reduced yet equivalent data [82, 83, 27]

3.8 Methods adapted in the dissertation research

Among various methods introduced in this chapter, this dissertation research adapted an indirect pairwise choice task to investigate the influence of musical selection on listeners' preference to multichannel reproduced piano sound. The multichannel piano sound fields were captured via four multichannel microphone techniques with their optimum placement for the best sound quality of each in the recording venue. For the pairwise test, 36 listeners' preference choice data were collected. Listeners' choice data were then transformed to the equivalent continuous scale values via a Matlab function developed by Wickelmaier and Schmid [64].

In order to parametrically analyze the listeners' preference, eight trained listeners elicited perceptual attributes that characterize the given stimuli using Triadic Comparison. This method presents three randomly chosen stimuli to the listeners and asks to choose the one stimulus most perceptually different from others. Subsequently, listeners are asked to generate an adjective to describe the way in which the stimulus chosen as the "odd" and another adjective to describe the way how the two other stimuli is similar. From the individually elicited descriptors, five terms

were selected based on the frequency of occurrence. The bipolar adjectives that anchored selected five terms and the definitions of terms were shown in the Appendix A. It is worthwhile to note that listeners' preference were estimated via direct scaling method; this dissertation research assumed that the trained listeners' affective response to multichannel sound field could represent the affective response of mother population as Olive *et al.* did in their study [84]. [1]

The magnitude of six attributes (five elicited attributes plus preference) were estimated for each of four versions of a single performance (musical selection), and these estimates were indicated by the listener adjusting a pointer along each of four slider bars on the GUI (please refer Figure 8–3) that coded the direct ratings with an internal resolution of 100 points between the two extremes. This multi-stimulus rating is similar to the MUSHRA (MUltiple Stimulus with Hidden Reference and Anchor) [85] method except that the method used in this research did not have a reference stimulus. Whereas the non-reference model (also referred to *intrusive* model) is inferior to the one with reference in their performance and prediction accuracy as Soren and Zachrov pointed out [28, Chap. 1], it was not practically possible to create a reference signal for an arbitrary multichannel piano sound of various periodic compositions.

[1] Please refer the Conclusion chapter 9.2 on the page 136 that has in-depth discussion on this topic.

Attributes	R	R^2	R^2 **Change**	F **Change**
Sharpness	.422	.178	.178	55.158
Sharpness, Width	.482	.232	.054	17.686
Sharpness, Width, Bass Tightness	.525	.275	.043	14.987

Table 3–1: Stepwise Multiple Regression results for predicting preference ratings from combinations of the five sets of attribute ratings. The table shows the change in the amount of variance (R^2) for which the model accounts as predictors are added to the model. Only statistically significant changes in R^2 are reported in this table (at a *Type I* error probability of $alpha = .01$).

These ratings were analyzed via PCA in order to investigate the interrelationship between six attributes and via Stepwise Multiple Regression Analysis to examine the attributes that has significantly accounted for the variance of the preference. The detailed results of statistical analysis have been reported in the series of publications[86][65]. One of examples is shown in Table 3–1 that shows the result of stepwise multiple regression analysis that examined the change in the amount of variance (R^2) of preference according to successful addition of independent variables, ratings of five salient attributes. Only three predictors - Sharpness, Width, and Bass Tightness - made significant progressive improvements in the prediction of preference according to the F Change values.

CHAPTER 4
Summaries of Publications

4.1 Publication I: An Examination of the Influence of Musical Selection on Listener Preferences for Multichannel Microphone Technique

Four solo piano pieces composed in the European concert music tradition and deemed to be approximately representative of various eras were selected and included works by: Bach, Schumann, Brahms, and a contemporary piece. The concept of genre or era in the history of Western composition is often hotly debated by musicologists. However, of interest in this study was not that a single composition might be taken to be symbolically representative of a specific epoche of composition. Rather, of interest was that the musical selections represent a wide variety of the acoustical possibilities regularly encountered by audio engineers and producers recording "classical" piano music. In short, different musical selections activate the piano-hall acoustical system differently. As a result, it was hypothesized that some microphone techniques might be preferred for certain musical selections within the hall in question.

Four surround microphone arrays were also selected: Fukada tree, Polyhymnia pentagon, OCT with Hamasaki square, and SoundField with surround decoder. The arrays were chosen considering the theoretical background of each array, the transducer and operational characteristics of the microphones required, the experience of the authors, and other practical issues. During the recording process each microphone array was optimally placed in the concert hall and each musical piece was

simultaneously recorded by all microphones. As stated in the previous section, the position of each array was varied in between musical selections if needed. All musical excerpts were performed in the same concert hall by a single musician, and played on a single piano.

After the recording, preference testing was conducted using male and female recording engineers as participants. All recordings were edited in order to have the same duration and processed for same loudness in multichannel reproduction with 5.0 playback. A two-alternative-forced-choice method (2AFC) was used during the listening experiment and listeners were asked to report a subjective preference between two randomly selected stimuli. The data from experiments were subsequently analyzed to build a preference model using statistical tools incorporated with an indirect scaling based on pairwise comparison.

Results show that for three of the musical selections, listeners on the whole were somewhat indifferent to whether the Polyhymnia Pentagon or Fukada Tree produced the best image (with both estimated preference scale values contained within each other's 95% confidence intervals). For the Brahms selection, on the other hand, the imagery associated with the Fukada Tree was highly favored. Even though no single microphone technique was chosen as most preferred throughout all musical selections, listener preferences with regard to microphone technique have been modulated by musical selection (an aggregate of the composition in question, the performance practice used by the musician, and the performance itself).

4.2 Publication II: Predicting Listener Preferences for Surround Microphone Technique through Binaural Signal Analysis of Loudspeaker-Reproduced Piano Performances

Previously four solo piano pieces were recorded using four different surround microphone techniques to produce a stimulus set of 16 items - four musical selections for each microphone technique. These stimuli were presented through a five channel full-range reproduction system for pairwise preference choices, and the results of that test could be described in terms of the interaction between program material and surround microphone technique.

In order to account for the dependence of preferences for surround microphone technique on the program material being presented, an attempt was made to predict the obtained preference choices on the basis of the reproduced sound signals alone. Therefore, a number of electroacoustic measures on the test stimuli were examined via stepwise multiple regression, with fit preference scale values as the criterion (dependent variable). These electroacoustic measures included standard binaural parameters determined for the reproduced soundfield, such as mid-side ratio, and interaural cross correlation (IACC).

In an attempt to develop a quantitative model for predicting listener preferences from physical measures of the stimuli, 16 potential predictors were submitted to stepwise multiple regression analysis, using as the dependent variable the preference scale value obtained from 36 listeners for 16 stimuli. The stepwise analysis showed that two predictors - `ESI`, ear signal incoherence and `SBR`, RMS ratio between high frequency and low frequency portion (with 250 Hz cutoff frequency) of side signal

RMS values, or side-bass ratio - accounted for more than 80% of the variance of log-scaled dependent variable (listener preference) for both of two independently tested groups of listeners.

$$\log_{10}(P_1) = 2.03 \cdot E - 0.21 \cdot S + 1.37$$
$$\log_{10}(P_2) = 2.33 \cdot E - 0.22 \cdot S + 1.21$$
$$(E = \texttt{ESI} \quad and \quad S = \texttt{SBR}) \tag{4.1}$$

The regression equations 4.1 derived for the two groups were practically identical despite the differences that exited in methods of data collection used for the two independently tested groups of listeners.

4.3 Publication III: Deriving Physical Predictors for Auditory Attribute Ratings Made in Response to Multichannel Music Reproductions

Five attribute scales were previously constructed on the basis of the results of a verbal elicitation task undertaken by the eight listeners who were engaged in Tonmeister training program at McGill University[86]. The attribute scales were anchored by five pairs of bipolar adjectives upon which the eight listeners reached some consensus, and included the following adjective pairs: Wide ↔ Narrow, Distant ↔ Close, Focused ↔ Diffused, Sharp ↔ Dull, and Tight-Bass ↔ Muddy-Bass. The same eight listeners made direct ratings of perceived magnitude associated with each attribute for the stimuli. Analysis on those ratings showed that two principal components accounted for most variance of five attributes. The first principal component was associated with attribute **WIDTH** while the second was with both attributes **BASS-TIGHTNESS** and **SHARPNESS**.

Unfortunately, those salient attributes did not show a strong correlation with the previously found physical measures, `ESI` and `SBD`, which caused a reason to examine the new physical measures associated with the attributes. For the physical measure of the attribute **WIDTH**, binaurally recorded signals were split to 25 Equivalent Rectangular Bands (ERB). Then each band's correlation between left-right ear signal was calculated based on the equation 4.2. SLi and SRi represent the Standardized Left and Right ear signals at ith ERB respectively. These correlations measured on each ERB were summed to a single number which represented a predicted magnitude of **WIDTH** on a given duration of a stimulus (w in the equation 4.3).

$$f(i) = \sum_{k=1}^{N} \frac{SLi(k) \cdot SRi(k)}{N-1} \tag{4.2}$$

$$w = \sum_{i=1}^{25} (1 - f(i)) \cdot g(i) \tag{4.3}$$

The prediction model of **BASS-TIGHTNESS** was implement as same way as **WIDTH** prediction model. The ear signal correlations up to the 9th EBR whose center frequency was 400Hz were summed to generate a simple metric for the predicted magnitude of **BASS-TIGHTNESS** (Bt in equation 4.4)

$$Bt = \sum_{i=1}^{9} (1 - f(i)) \cdot g(i) \tag{4.4}$$

It would be worth to note that the ERB weighting function $g(i)$ in the two equations 4.3 and 4.4 was constant as 1 in this study. Further investigation of $g(i)$ with regard to the physical measures of **WIDTH** and **BASS-TIGHTNESS** might result better prediction. Later preference ratings were related to the two sets of physical measures via stepwise regression and about 70% of total variation in preference ratings was accounted for by the two measures. These physical measures could serve as a prediction model for listeners' affective response (such as preference) of an arbitrary multichannel piano sound.

CHAPTER 5
An Examination of the Influence of Musical Selection on Listener Preferences for Multichannel Microphone Technique

This chapter contains the exact the copy of the paper published in the proceedings of the 28^{th} International AES conference held at Piteå, Sweden in June 2006, which was written in collaboration with four co-authors - Martha DeFrancisco, Kent Walker, Atsushi Marui, and William L. Martens.

ABSTRACT

Four solo piano pieces performed in the European concert music tradition and deemed to be representative of differing stylistic periods were recorded using four surround microphone arrays: Polyhymnia Pentagon, Fukada Tree, Optimized Cardioid Triangle with Hamasaki Square, and SoundField MKV (5-channel processing via SP451). Each array was positioned and balanced in order to optimize its perceived sound quality as opposed to being positioned solely according to theory. Blind preference testing was subsequently conducted using audio engineers and musicians who auditioned the recordings through a full-range five-channel reproduction system compliant with ITU BS.775-1. The results show that listener preferences for multichannel microphone techniques may be influenced by musical selection (a particular interpretation/performance of a given composition).

5.1 INTRODUCTION

5.1.1 Taste: The Spectacles Worn by Reason?

> "Of all the natural gifts taste is the most easy to recognize and the most difficult to explain. It would not be what it is if it could be defined, since it judges matters that are not in fact capable of being judged, serving - if such an idea is permissible - as a pair of spectacles to reason... Everyone has his own individual taste by which he establishes his own private scale of values among the things that appear to him good and beautiful." - Jean-Jacques Rousseau (1712-1778) [87, Chap. 2]

As pointed out by Pierre Boulez through his use of these quotations in his lecture "Taste: The Spectacles Worn by Reason?" [87], not much has changed in aesthetics since the time of Rousseau. "Revolutions in taste" are clearly quantifiable within history, yet we are only just beginning to understand the nature of taste itself. At the same time, according to Boulez, within academic and educated circles that study music, there is a disinclination to discuss taste: "Is [taste] never mentioned because people think of it as a natural, familiar gift whose existence there is no point in admitting, or as a disgraceful disease to be discussed only in vague terms and behind closed doors?"

Boulez's comments and questions, although most specifically aimed at European concert music of the 20^{th} century, have implications for the art of Sound Recording. Taste *per se* is not necessarily understood and is not often addressed in recording literature, yet it seems that it plays some role in the production process, at least

according to a comparison of microphone arrays conducted by the ORF (Austrian Broadcasting Corporation) [10] and the IRT (Institut für Rundfunktechnik) [88]:

> "The individual recordings were mixed and optimized ***according to taste*** by the relevant protagonists if present (Theile/Wittek for OCT and Gernemann for Stereo+C); The other systems were adjusted by participants with specific experience regarding one or the other system." [88]

Indeed, this process of optimization of microphone arrays is well known to the *Tonmeister*, a specialist who may be defined as a recording engineer who combines specific technical training with solid musical studies [89]. According to this online definition taken from the German-language Wikipedia, the Tonmeister mediates between the artistic demands of the performer and the technical realization of the recorded sound. The question here that must be asked about Tonmeisters, is that regarding whether this optimization process is done strictly according to personal ***taste***, or whether Tonmeisters rely upon an objective standard that is based upon some consensus derived from their training.

It has been said that Tonmeisters participate in a discipline which may be described primarily as ***re-creative***, meaning that its aesthetic aim is the re-creation of natural acoustical properties of original sound sources in recording spaces. Tonmeisters, therefore, are not free to make optimizations according to any whim or fancy; instead they employ a skill of representation built over years of experience, aural training, and natural observation. However, if this ability were absolutely consistent between all Tonmeisters, every one of their recordings would sound the same, or at least alike given similar acoustical conditions and techniques. While there is a degree

of consensus among renowned record labels of Concert Music (i.e. Western Music) with regards to sound quality, there are also discrepancies. Thus enters a semantic conundrum: What word should one use to discuss differences in Tonmeister judgment as applied to the recording process?

The last 25 years have seen an emerging scientific literature that has examined the perceptual qualities of reproduced sound. These investigations have euphemized and sanitized *taste* in terms of simple dominance relations between different versions of a given reproduction. In effect, the complexity of comparisons between auditory imagery is reduced to a choice of one version over another, and thus dominance is operationally defined in subjective evaluation in terms of *preference*. In the case of Tonmeister judgment, this term may be no better; naive listeners also have preferences. Regardless of the special skills of the Tonmeister, denying the use of taste, however minimal in the recording process, denies the humanity of the Tonmeister. As stated by Rousseau, taste is a natural gift; and who can evaluate reproduced sound quality without employing it?

A similar semantic argument has emerged with regards to *naturalness* [6]. When discussing the creation of natural recordings, the most often asked question seems to be "natural to whom?" Rumsey [6] describes the struggle for naturalness as a compromise between aesthetic practice and microphone theory: "...optimization by the sound engineer, will be the better, the more flexible the stereophonic recording technique is [6]." While it is true that there is "nothing more practical than a good theory" [10], ultimately it is the Tonmeister who must consider all possible variables. Despite the acknowledgement that optimization plays a critical role in the creation

of recordings, it seems its details have gone largely undocumented in academic Sound Recording literature. It was, therefore, one of the aims of this study to document in detail the processes of optimization used during the creation of multiple version of a recording for controlled comparisons, in the hopes of not only learning something about listener preference, but also of discovering and developing for future use additional multichannel microphone techniques that might be more flexible. The optimization process was largely directed by the second author, who has more than thirty years of experience as a Tonmeister and producer. Tonmeisters often optimize microphone arrays through slight alterations of placement. By documenting herein the asymmetries of placement introduced during the recording process, the authors do not mean to suggest that the original configurations are in anyway deficient in theory. The objective of these alterations was simply to realize the best potential of each array in question, while also providing grounds for comparisons between techniques to be done in the most fair manner.

This study implements optimized multichannel microphone arrays within the context of an investigation of preference. The objective was not to show that some techniques are simply more preferred than others, but rather to examine the relationship between preference for microphone techniques and musical selection, which includes here both a composition and its interpretation or performance. This relationship has been suggested by other authors, however little data has been collected. The current investigation thus attempted to address the following primary question:

What influence does the reproduced musical selection (combining both a composition and its performance) have on listener preferences

for the auditory imagery created by different multichannel microphone arrays?

5.1.2 Existing Published Multichannel Microphone Comparisons

The development and study of microphone technique for purposes of sound recording constitutes a significant portion of the literature generated by organizations such as the Audio Engineering Society. In particular, there has been much material generated with regards to stereophonic techniques [90], binaural recording [91], and multichannel sound [92]. Recent studies have concentrated on 3-2 recording and reproduction, also called 5.1 ITU-standard surround [4], which is also the focus here. Despite the significant body of material generated, there have been relatively few *published* comparison studies incorporating *both* a wide variety of simultaneously recorded arrays and formal collection of subjective evaluation data from listeners. The logistics involved in creating such comparisons are quite formidable. Obtaining large numbers of high-quality microphones, patient yet qualified musicians, and appropriate facilities is difficult. It is also time-consuming to construct and implement the tools required for proper subjective testing. It is not surprising, therefore, that only two English-language studies somewhat similar to the effort published here could be located by the authors[1] .

In the previously mentioned study conducted by the ORF, Camerer and Sodl [10] evaluated the performance of seven different microphone arrays that they describe

[1] Failed attempts were made to locate [93], a thesis cited in [11].

as belonging to one of four categories: the *No Sweet-Spot Group* consisting of techniques involving curtains of microphones; the *Sweet-Spot Group* consisting of univalent microphone techniques incorporating time-amplitude trading; the *Natural-Illusion Group* consisting of techniques incorporating solid psychoacoustic theory; and the *Verisimilitude Group* concerned with wavefront reconstruction (e.g., using Ambisonics). The seven microphone techniques evaluated in their study were the following: Stereo+C and Hamasaki Square, Decca-Tree and Hamasaki Square, Optimized Cardioid Triangle (OCT) and Hamasaki Square, Ideal Cardioid Arrangement (INA), Schoeps KFM 360, OCT Surround, and SoundField MKV + Processor SP451. Simultaneous recordings were made of the Radio Symphony Orchestra of the ORF (RSO Vienna) performing in Austrian Radio's Grosser Sendesaal (large broadcast hall). Compositions consisted of 2 musical selections (Mozart and Berio) of 2 minutes each, played-back consecutively. Subjective ratings were collected from 18 subjects with regards to the following attributes: spatial presentation of the orchestra (wide-narrow, close-distant, deep-flat, stable-unstable, precise-blurred); timbre (satisfactory-unsatisfactory); and spatial imaging (perfect versus imperfect spatial impression, too much indirect sound, surround channels identifiable).

In introducing a report of another similar study Kassier et al. [94] remarked (after Rumsey [6]) that there is a lack of test materials for comparison of surround-sound microphone techniques. They also noted that the most common techniques generally fall into two groups: those that use five-channel main microphone techniques (univalent design) and those that use techniques with front-rear separation.

Subsequently, Kassier et al. created simultaneous recordings of eight different techniques, *all* of which incorporated front-rear separation; a distance of seven meters was chosen between the front and rear portions of all arrays. Front triplet arrays included: Fukada-like Tree, OCT-Inspired Technique, INA-3 technique, and a near-coincident Klepko-inspired technique. Rear arrays included: IRT-Cross, Hamasaki Square, Klepko-inspired technique, and a spaced cardioid array. A variety of small ensembles and speech were recorded, and subsequent subjective ratings of listener preference were collected from six experienced participants using a proprietary listening test software (ALEX). Note that this was not designed as a blind preference test, as all of the items were clearly marked (coded) and were not randomized in their presentation.

5.2 RECORDINGS

5.2.1 The Musical Selections

Four compositions for solo piano judged to be representative of different stylistic periods were selected by the pianist, Professor Thomas Plaunt:

- *Late Baroque*: Johann Sebastian Bach (*b* Eisenach Germany, 21 March 1685; *d* Leipzig, 28 July 1750), "Variation 13", *Goldberg Variations* (BWV 988, *circa* 1741)
- *Early Romantic*: Franz Peter Schubert (*b* Vienna Austria, 31 January 1797; *d* Vienna, 19 November 1828, "1. Allegro ma non troppo", *Sonata in A minor* (D537, 1817)
- *Romantic*: Johannes Brahms (*b* Hamburg Germany, 7 May 1833; *d* Vienna Austria, 3 April 1897), *Ballade in D minor* (op. 10 no. 1, 1854)

Frequency	63Hz	125Hz	250Hz	500Hz	1kHz	2kHz	4kHz
RT_{60}	2.3 sec	2.0 sec	1.7 sec	1.8 sec	1.8 sec	1.7 sec	1.4 sec

Table 5–1: The reverberation time RT_{60} of the Pollack hall in each frequency band from 63 Hz to 4 kHz

- *Contemporary Improvisation*: Thomas C. Plaunt, *untitled* (2006)[2]

5.2.2 The Piano and The Hall

The piano used in this experiment was a New York Steinway model Concert D. All excerpts were recorded in the Schulich School of Music's Pollack Hall (McGill University). The dimension of this hall is 36 m long, 18 m wide and 12 m high, with a 590-seat capacity. The reverberation time (RT_{60}) of Pollack hall when empty may be seen in Table 5–1. Acoustical curtains were additionally used to reduce the reverb time at higher frequencies.

5.2.3 The Microphone Arrays

The pre-optimized techniques used bare close resemblance to their namesakes to the best knowledge of the authors. Although all of the arrays were first positioned according to original recommendations, inter-microphone spacings and angles were slightly altered according to the requirements of the recording session, thus optimizing the arrays in terms of perceived sound quality. Photographs of the implemented arrays may be seen in Figure 5–1. Although a binaural recording was made, and a

[2] Some readers may not be familiar with improvisation in the context of European Concert Music (Western Music). For varying perspectives on its history, please consult [95] and [96].

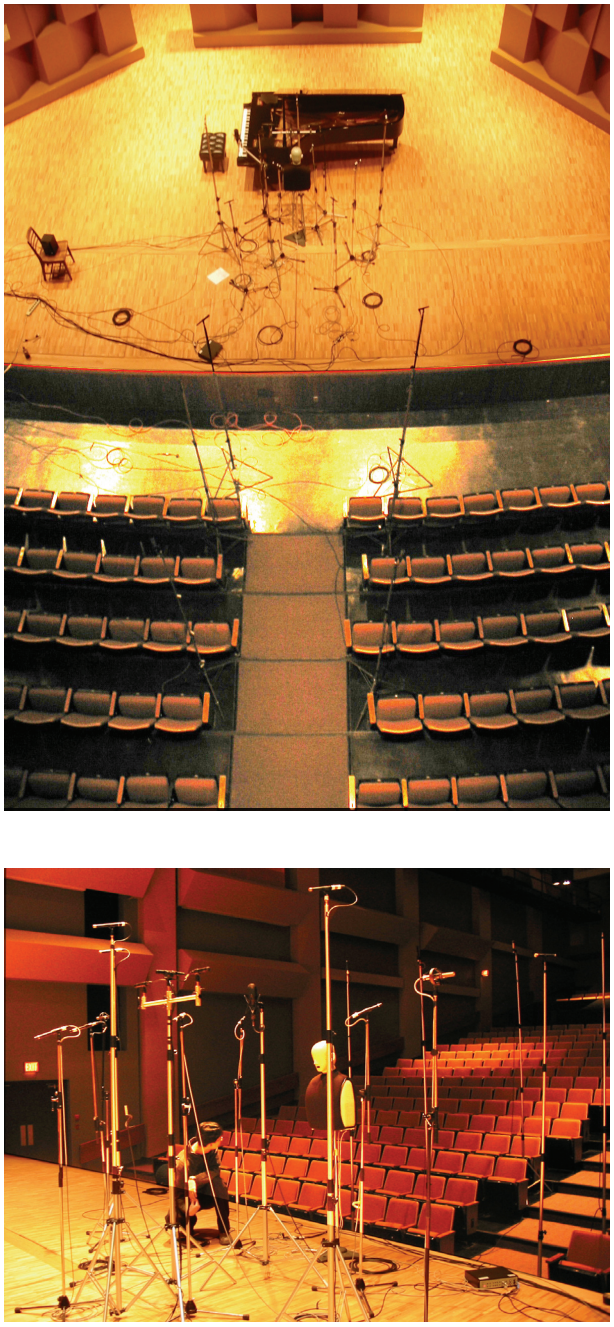


Figure 5–1: Placement of four microphone arrays in Pollack hall, McGill University. (A) Top view; vs. (B) Close view

dummy-head is visible in Figure 5–1, it was not used for preference testing in this study.

Fukada Tree

Five-channel univalent microphone techniques incorporating time-amplitude trading (i.e. arrays of directional microphones where there is a one-to-one microphone-speaker ratio) have been discussed in the literature of the AES since at least 1991 [97]. However, the term Williams Tree (after Michael Williams) is not used here as the Williams Curves were not adhered to and much larger inter-microphone spacings were employed. Also, the rationale for microphone placement, the adjustment of angles and spacings according to experience and careful listening by the Tonmeister rather than by theoretical basis alone, was closer to techniques suggested by NHK broadcast engineer Akira Fukada [11]. The implemented technique, consisting of five DPA type 4011 microphones may be seen in Figure 5–6. Readers will note that the array is not quite symmetrical; the decision to displace microphones in this way was made because the original symmetrical array sounded unbalanced for this application. This is due in part to the radiation pattern of this instrument. During the session outrigger pressure microphones were also initially considered (as suggested by Fukada for orchestral recording), however, it was decided that they were not needed.

Polyhymnia Pentagon

An array of five omnidirectional pressure microphones using large spacings was also implemented. This technique was originated by Polyhymnia International (formerly the Philips Classics Recording Department). The technique specifies that the microphones be arranged in a circle with a radius of approximately 3 meters,

placed at the same angles relative to the median plane as the loudspeakers in ITU-Recommendation BS.775-1 [98] [99]. This technique is often described as a multichannel version of the Decca-Tree, a popular stereophonic array used in music recording and film sound. While both techniques do implement widely-spaced pressure microphones, they are actually quite different. The Decca-Tree typically uses much smaller spacings between microphones, approximately 1.5 m to 2.5 m between left and right, and does not use an equidistant or circular radius (the center microphone is typically 0.8 m to 1.2 m in front of the left and right microphones) [100]. The Decca-Tree was also originally designed for Neumann type M50 microphones, which are directional at frequencies starting at approximately 6 kHz, roughly one octave lower that the directional characteristics of many pencil-style pressure transducers. The Polyhymnia Pentagon implemented in this study may be seen in Figure 5–7. It should be noted that an equidistant radius was abandoned. As with the Fukada Tree, asymmetrical placement was chosen in order to balance the frontal sound image. The angles of pencil-style pressure microphones are important to the quality of recorded sound, as for high-frequencies (above 10 kHz or so) there will be inter-channel level differences as well as inter-channel time differences, therefore, time was taken to adjust angles in the vertical plane in order to achieve the desired amount of brightness. Horizontal distances and angles are shown for the DPA type 4003 and 4006 microphones used in Figure 5–7. The front triplet of microphones was directed forwards, and the rear-facing couple was directed towards the back of the hall.

OCT with Hamasaki Square

The Optimized Cardioid Triangle (OCT) was first proposed by Günther Theile [101]. This method is known to reduce crosstalk between channels by incorporating two hyper-cardioid microphones facing $\pm 90^\circ$ from the instrument. As shown in Figure 5–8, this array consists of a center cardioid microphone placed 8 cm forward, and right and left hyper-cardioids with an adjusted spacing ranging from 40 cm to 90 cm depending on the intended recording angle. During the recording session the distance between the left and right microphones was adjusted in order to achieve the best possible frontal image. Optional low-passed pressure microphones may also be used for enhanced low-frequency response. Signals from these optional microphones (coincident in this case to the left and right hyper-cardioids) were low-pass filtered and summed with the high-passed filtered portions of the left and right microphones at 100 Hz. It is possible to combine this frontal array with several rear techniques such as: OCT surround, IRT cross [6], or Hamasaki Square [16]. Among them, Hamasaki Square was chosen for this study. Positive lobes of four bi-directional microphones are pointed towards the walls of the acoustic space in question. These microphones are routed to left, surround left, right, and surround right channels respectively. This technique thus encodes diffuse field ambience and sidewall reflections, as direct components are minimized by aiming the null of bi-directional microphones at the source. A diagram showing the distances and placement of each microphone type is shown in Figure 5–8.

SoundField MKV + Processor SP451

Ambisonics is a well-known technique which captures a sound field at a single point by encoding sounds from all dimensions in terms of pressure and velocity components (known as B-format). Encoded B-format is the equivalent of four coincident microphones: three purely pressure-gradient microphones (velocity microphones with regards to operational principle) facing forwards, sideways, and up and down respectively; as well as one pressure microphone. These components may then be subsequently decoded to any required loudspeaker configuration, and may also be matrixed to create the mathematical equivalent of any desired coincident microphone array, pointing in any desired direction, with any desired inter-microphone angle [102]. The most convenient method of capturing an Ambisonic signal is to use a SoundField microphone, which is composed of a tetrahedral sub-cardioid capsule array whose signal is known as A-format, which is then transformed to a B-format equivalent. The SoundField type MKV was used in this study. Its B-format output was processed by the SoundField model SP451 surround processor, which generated discrete 5.1-channel outputs. The decoding process most properly used to transform B-format into 5.1 is called a *Vienna* decoder and the resulting speaker feeds are termed G-Format [88]. Ideally, the SP451 should offer the user the ability to produce a 5-channel full-range signal. However, this was not possible, and the 0.1 channel was added in equal amounts to L,C,R,LS,and RS. The position of this microphone was adjusted as desired, and the final array is shown in Figure 5–9.

5.2.4 The Recording System

With the exception of the DPA type 4003 microphones used in the Polyhymnia Pentagon (routed directly to a Millennia Media model HV-3D preamplifier) and the SoundField MKV, all microphones were preamplified by a GRACE Design model 802. All line level signals were subsequently converted and recorded using a ProTools HD system at 192kHz / 24bits.

5.2.5 Mixing

Recordings were mixed in order to provide an even-ground for comparison. Loudness mismatches between recordings and some inter-channel level differences within recordings were manipulated. In particular, surrounds were slightly reduced to a point where they provided the desired surround effect (approximately -3 dB). A limited amount of equalization was also applied to the surround channels of the Polyhymnia Pentagon (a high-pass filter at 80 Hz) in order reduce some of the perceived low-frequency build-up. Mixed sound files were down-sampled to 48 kHz at 24 bits and saved as .wav files.

5.3 LISTENING EXPERIMENTS

5.3.1 Subjects

A total of 36 listeners took part in the listening experiments. 26 were either professional recording engineers or sound recording students. The other participants were music students and music faculty members. Age varied from 20 to 47 years, and none of the listeners reported having any hearing disorder.

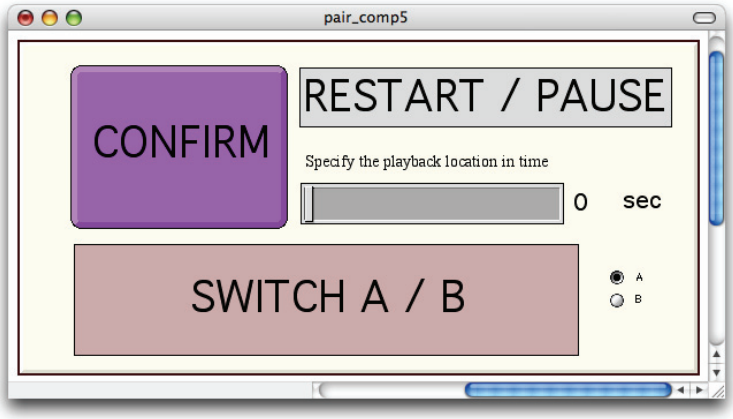


Figure 5–2: Graphic User Interface (GUI) implemented by MAX/MSP to present stimuli and collect preference choice data

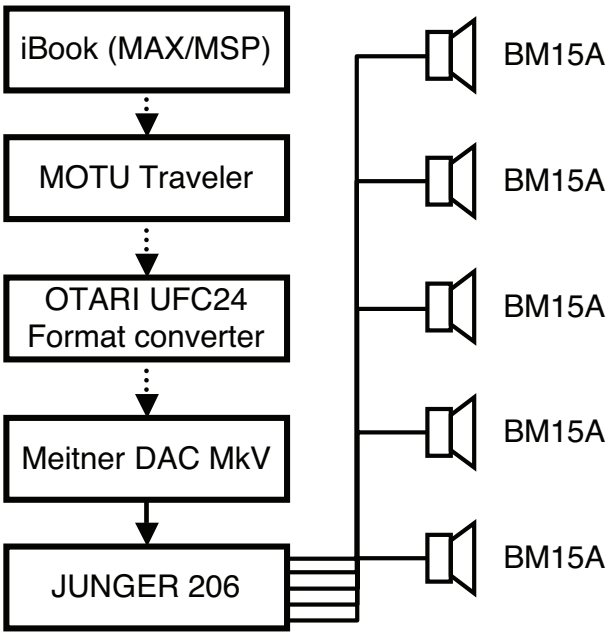


Figure 5–3: Signal path for the reproduction of multichannel audio. Dotted lines represent digital signals, while solid lines represent analog

5.3.2 Reproduction System and User Interface

Five active full bandwidth loudspeakers (Dynaudio model BM15A) were placed in MARLAB (**M**ultichannel **A**udio **R**esearch **LAB**oratory) of McGill University according to ITU-Recommendation BS.775-1 [4], with a height of 1.2 m from the floor and with a radius of 1.5 m. The ambient noise in the room was 27dBA measured from the central listening position. Calibration of the loudspeaker levels was performed using a Brüel & Kjær type 2235 sound level meter with A-weighting and fast response. Each speaker output was individually calibrated to 78dB SPL using a -18dBFS pink noise input signal, giving 85dB SPL in total for all five speakers.

A customized MAX/MSP (Cycling 74) Graphical User Interface (GUI) was created in order to collect preference data. This interface allowed listeners to select between two multichannel recordings in real time. As Figure 5–2 shows, listeners were given the opportunity to break and re-start at any time. Listeners were also asked to consider each presented pair as if it were a brand-new comparison (i.e. they were told not to worry about previous responses). Preference choices were simply indicated by pressing the *CONFIRM* button at any time during playback. Each successive pairwise comparison appeared and began playback automatically. All sound files were 48kHz and 24bits, digitally reproduced through a Mark Of The Unicorn Traveler Firewire audio interface, passed through an OTARI model UFC 24 digital Universal Format Converter, and converted to an analogue signal by a Meitner type DAC MkV. Analogue distribution and level control was accomplished with a Junger

model 206. The detailed signal path used in the listening experiments is illustrated in Figure 5–3.

5.3.3 Preference Choice Testing

Each listening experiment was generally conducted in the following manner: starting with an informal listening session, listeners were allowed to hear all four versions of the four piano performances; subsequently the main session was conducted, where each listener completed four blocks of binary paired comparisons. Only two sound files were presented at a time, and subjects were asked to choose the preferred one. A context for preference was provided: listeners were instructed to select the file that they would rather listen to for an extended period of time in their home over loud-speakers. While direct scaling procedures are often used in audio-related research projects, binary paired comparisons were used here because, as explained in [63], indirect scaling using pairwise judgment does not rely on implicit and untested assumptions, and may be used to reveal relatively small differences existing among stimuli. During the experiments participants were not told which microphone arrays they were hearing and were simply presented with two stimuli labeled A and B. Listeners compared each pair twice but in opposite order, creating listening sessions comprised of a total of 48 pairwise choices (four blocks of 12 preference choices).

Out of 36 listeners, two groups of 18 listeners were formed which completed trials according to different trial ordering schemes. Pairwise-comparison trials themselves were otherwise identical (as described above). For the first group, all trials for a given musical selection were completed in a single block. In other words, each

block consisted of randomized microphone arrays presented for each musical selection. This approach to trial ordering has been termed *successive-treatment design* [103]. The second group also completed four blocks of 12 preference-choice trials, but musical selection was randomly assigned from trial to trial. This approach to trial ordering has been termed *intermixed-treatment design* [103]. More discussion of the experimental design employed here, and more detailed analysis of the contextual effect observed as a result, are reported in a companion paper (in this proceedings [65] [3]).

5.4 RESULTS

Collected preference scores of each group were processed using a MATLAB function which estimates choice model parameters from paired-comparison data [64]. The derived preference scale values show the relative merit of each stimulus in comparison to the others presented. These values do not indicate the degree of preference among microphone techniques *per se*, but rather the relative preference for each stimulus within the given set. Figure 5–4 displays the results for the two most highly preferred microphone techniques in order to summarize how the estimated preference scales are modulated by musical selection when analysis is based upon combined preference choice data from all 36 listeners. For three of the musical selections, listeners on the whole were somewhat indifferent to whether the Polyhymnia Pentagon or Fukada Tree produced the best image (with both estimated preference scale values

[3] The contextual effect of the experimental design is summarized in section 8.2 on page 122 of this dissertation.

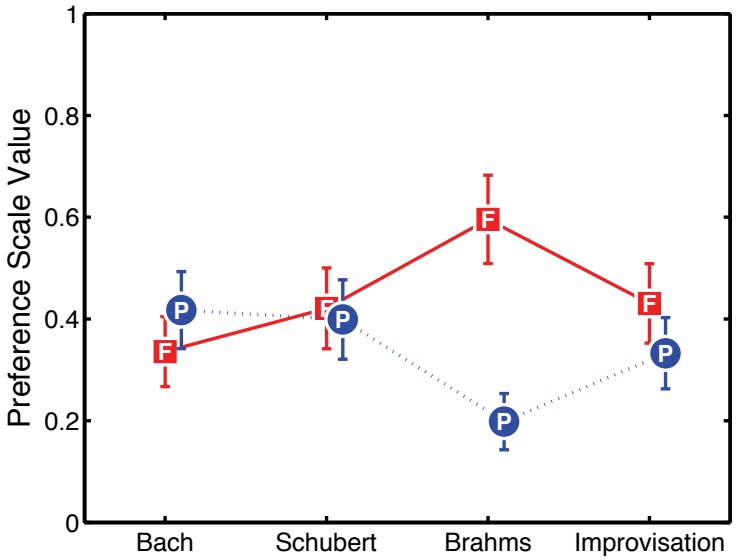


Figure 5–4: Preference scale values estimated from preference choice data collected from 36 listeners with regard to imagery associated with four piano pieces resulting from multichannel reproductions based upon two of the four microphone techniques included in the test. The two microphone techniques compared here are Fukada Tree and Polyhymnia Pentagon (plotted as F and P respectively). Data were pooled across two groups of 18 listeners, all of whom heard the pairs of stimuli in differing orders, but with intermixed versus successive trial ordering schemes (see text). Error-bars represent corresponding 95% confidence intervals.

contained within each other's 95% confidence intervals). For the Brahms selection, on the other hand, the imagery associated with the Fukada Tree was highly favored. Figure 5–5 shows the estimated preference scale values in all conditions separately for both groups of listeners. The bars were color-coded for microphone techniques as follows: Fukada Tree as indigo, Polyhymnia Pentagon as sky blue, OCT with Hamasaki Square as yellow, and SoundField as brown. The goodness of fit of the estimation model parameters was evaluated and shows that musical selection affects preferences for multichannel microphone techniques under certain conditions. More details on the statistical analysis of the obtained data, as well as an investigation of contextual effects observed here, will be reported in a companion paper [65]. Suffice it to say that listener preferences with regard to microphone array are more strongly modulated by musical selection in the intermixed-treatment group, as may be seen in Figure 5–5. No single microphone technique was chosen as most preferred throughout all musical selections.

5.5 DISCUSSION AND FUTURE WORK

Results clearly show that the Fukada Tree is greatly preferred for the Brahms musical selection, however, for other musical selections either the Fukada and Polyhymnia techniques could be chosen with roughly equal likelihood. Further investigations are underway in an attempt to determine the factors underlying these observations. Studies employing the same stimulus set are currently focussed upon the following: analysis of the physical signals, ratings of distinct perceptual attributes associated with imagery for all 16 stimuli, and replications of the preference tests in a variety of listening rooms. It does seem likely, however, that the derived preference

scales were influenced by the density of the Brahms composition in lower registers, which affected both the timbre and spatial projection of the piano recorded in the hall. It is equally possible that recordings of music from the Romantic stylistic period involve different aesthetic concerns. The preference scales derived here may nevertheless be compared to results from other studies presenting different musical selections and microphone techniques.

Kassier et al. [94] investigated general preference and showed that the Fukada Tree was *always* preferred for a variety of musical sources and speech. It is important to note that their Fukada Tree varied greatly from the tree employed here, as a very large degree of front-back separation (7 meters) was used. The recommended spacing by Fukada is less than 2 meters. Kassier et al. also used greater separations than recommended for the Klepko Method [12] and INA techniques. There is a trend in the literature which supports the use of separation between front and back elements of multichannel arrays, however, most techniques do not specify spacings as large as 7 meters [12] [104] [105] [106] [13]. Kassier et al. also did not choose to optimize the placement of microphones using Tonmeister judgment. Rather, they placed the center of all front and rear arrays at the same points in the room.

Camerer and Sodl [10] were primarily concerned with which techniques might sound best and, therefore, be adopted by broadcasting organizations for transmission of orchestra performances. The current study held two techniques in common their investigation: OCT with Hamasaki Square, and SoundField MKV with an SP451 processor. Subjective ratings of perceptual attributes collected by Camerer and Sodl for the SoundField system generally tended towards the negative end of the scale. On

the surface, their results appear to be similar with those published here, as derived preference scales were low for this system regardless of musical selection. Camerer and Sodl explain that the SoundField MKV combined with the SP451 produced a narrower sound stage and a "close and flat feel". Although the authors did not collect related attribute ratings in the current study, similar remarks were made by Tonmeisters during the recording process. Camerer and Sodl further state that,"It is not quite clear, why the SoundField system which can produce stunning results in 2-channel-stereo falls somewhat behind most of the systems...". The third author has had similar experiences with this microphone in stereo and the results published here speak for themselves. It seems possible that a better transcoding processor into G-format might produce higher sound quality (perhaps a process closer to a *real* Vienna decoder than the SP451 [10]). However, it also appears to be likely that inter-channel time differences (as well as level differences) are important to the quality of recorded sound in 3-2 systems. There seems to be growing consensus in the literature that low inter-channel correlation is important with regards to the reverberant field of reproduced sound [107] and that high-correlation produces undesired timbral effects in multichannel sound [104] [105] [108]. Decorrelation between front and rear components of arrays is additionally held to be important by the creators of several multichannel recording techniques. Some describe the presence of direct sound in the rear channels as disturbing. It seems, therefore, that while coincident arrays are capable of producing great results in 2-channel stereo, they may have limited success in multichannel recording and reproduction when not used

in combination with other techniques. What constitutes the basis for what information should be presented from the surround channels is still under discussion within the Sound Recording community, where terms such as spaciousness, reverberation, depth, and envelopement are often used in explaining the function of these channels. Regarding the use of these terms in this context, Camerer [10] writes:

> "These terms sound similar, but they mean quite different things. The author doesn̂t want to go into great detail here, these and other attributes are very much under discussion among the surround sound community. This lively dispute has led to several microphone systems according to (very often) taste, plain theory or pure guessing."

5.6 CONCLUSIONS

Estimated scale values published here show that musical selection can significantly influence preference for these techniques. Additionally, it has been shown that such estimations may be generated by binary paired comparisons of multichannel microphone arrays, a methodology which may help to reveal the details of preference. In order for preference of multichannel microphone arrays to be most accurately understood, the complex variables associated with musical selection should be further investigated. In addition, optimization of multichannel microphone arrays is critical to perceived sound quality. This tradition of slightly altering theoretical placement and angles has been long-practiced by Tonmeisters, and is documented here in order that the literature may more accurately reflect recording practice. By documenting these placements it is possible that new theoretical arrays may emerge.

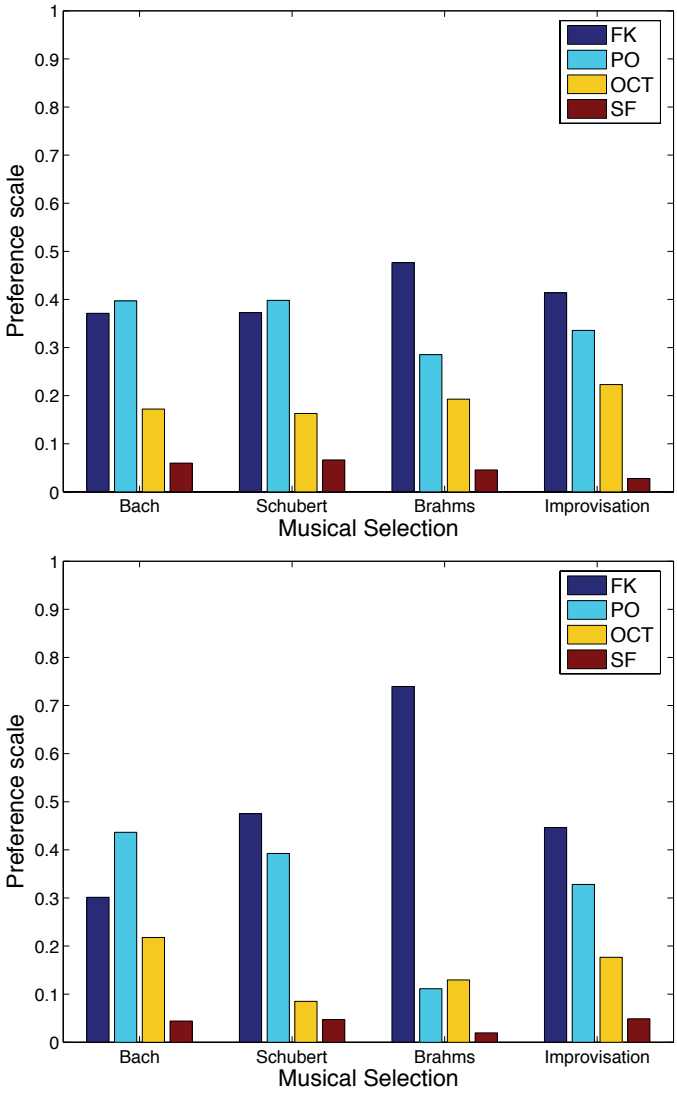


Figure 5–5: Preference scale values estimated from preference choice data obtained separately from each group of 18 listeners: Results given *successive treatment* are shown in the top graph; Results given *intermixed treatment* are shown in the bottom graph. The various musical selections are placed along the abscissa, and the bar of different color codes represent microphone techniques used in this study: Fukada Tree (FK) as indigo, Polyhymnia Pentagon (PO) as sky-blue, Optimized Cardioid Triangle (OCT) with Hamasaki Square as yellow, and SoundField (SF) as brown.

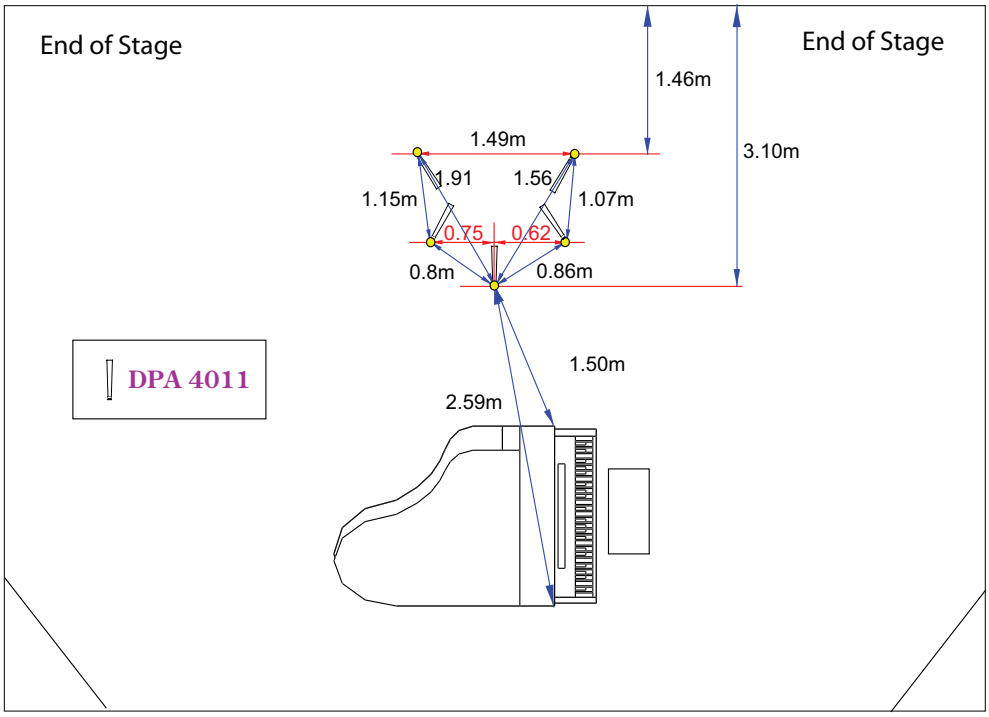


Figure 5–6: Microphone placement for the implemented Fukada Tree

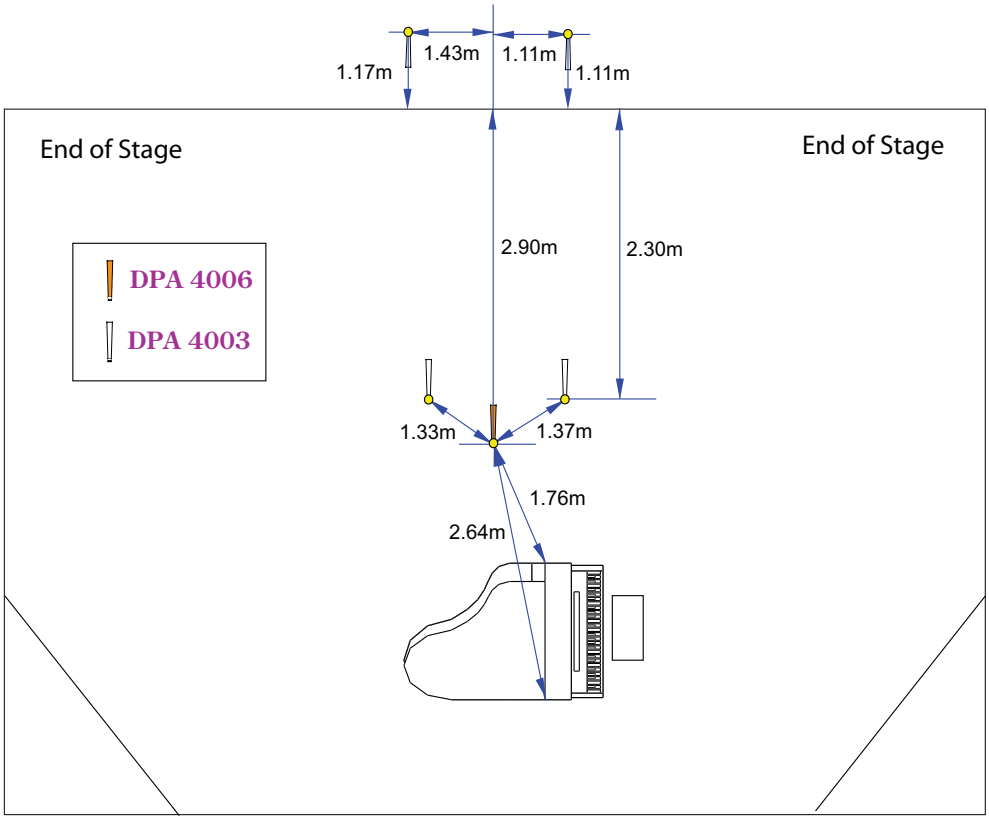


Figure 5–7: Microphone placement for the implemented Polyhymnia Pentagon

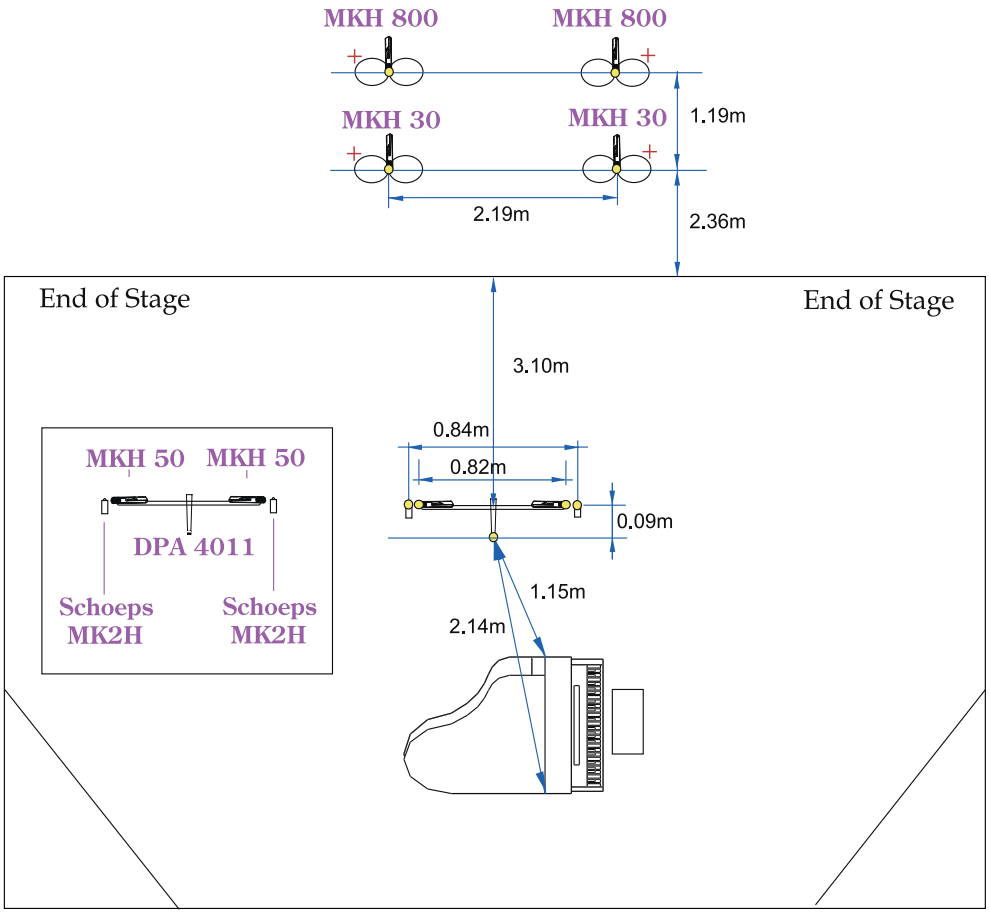


Figure 5–8: Microphone placement for the implemented OCT with Hamasaki Square

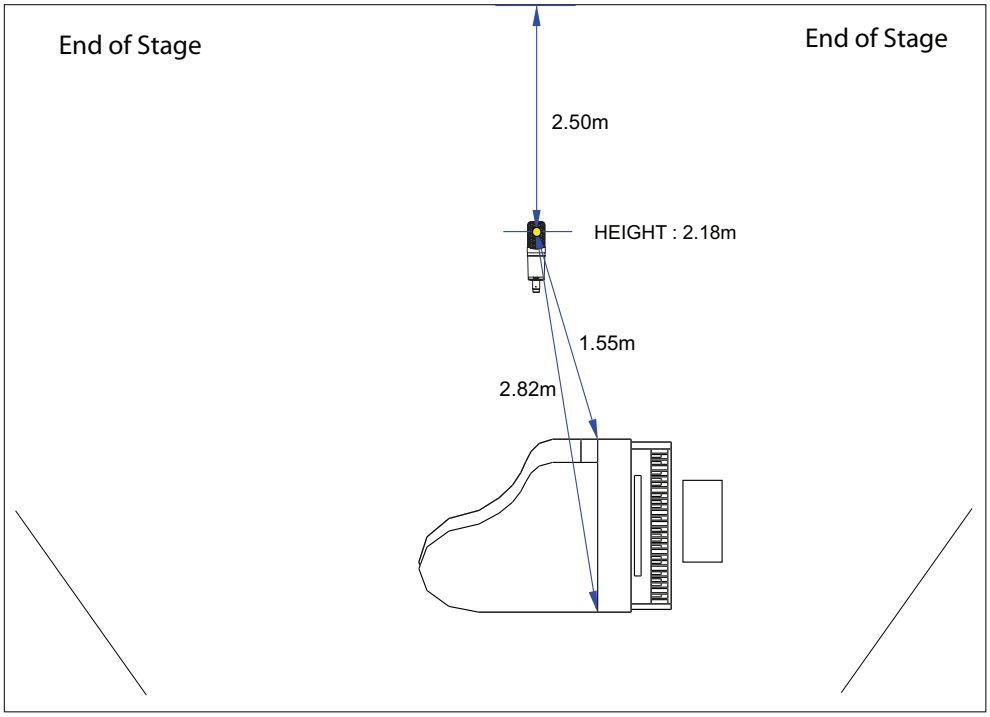


Figure 5–9: Placement of the SoundField MKV

CHAPTER 6
Predicting Listener Preferences for Surround Microphone Technique through Binaural Signal Analysis of Loudspeaker-Reproduced Piano Performances

This chapter contains the exact the copy of the paper published in the proceedings of the 121^{st} International AES convention held at San Francisco, USA in October 2006, which written in collaboration with two co-authors - William L. Martens and Atsushi Marui.

ABSTRACT

Four solo piano pieces were presented through a five-channel loudspeaker reproduction system for a pairwise preference test in a previous study, and the results of that test were described in terms of the interaction between program material and surround microphone technique. In an attempt to predict the obtained preference choices on the basis of the binaural signals recorded during loudspeaker reproduction of differing versions of these musical programs, a number of electroacoustic measures on the test stimuli were examined via stepwise multiple regression. The most successful prediction resulted from a combination of Ear Signal Incoherence (ESI) and Side Bass Ratio (SBR), regardless of methodological differences between two independently tested groups of listeners.

6.1 INTRODUCTION

The problem of how to assess the quality of spatial sound reproduction systems recently has been receiving increasing attention, which this year culminated in an AES workshop on "Spatial audio and sensory evaluation techniques" (see [109]). One of the key components required for solving this problem as identified there in an opening address by Rumsey [110] was the development of physical measurements that could be used to predict the perceived quality of reproduced sound fields. One of the primary difficulties that must be addressed in this endeavor is that listener preference reports are subject to biases due to contextual effects of various sorts [111]. For example, when the order in which stimuli are presented makes a significant contribution to the pattern of preference choices in a given experiment, then response prediction based solely upon stimulus parameters measured within an isolated stimulus presentation may as a result be less successful [112]. Since multiple musical test programs are typically presented within listener preference tests for a variety of spatial sound reproduction systems, stimulus presentation order is always a concern in these studies; and indeed, such a methodological manipulation was shown to have a significant effect upon the results found for two independently tested groups of listeners in a recent investigation of contextual dependency in a pairwise preference choice task [65].

In another recently published paper [113], the authors reported on the influence of changes in musical test program on listener preferences for surround microphone technique. The current paper addresses an issue that was left unexamined in that previous paper, which analyzed only preference choice data that were collected from

36 listeners. The unexamined issue concerns how the observed preferences for surround microphone techniques might be related to physical (instrumental) measures made on the multichannel loudspeaker stimuli presented to the listeners (cf. [114]). It may be that such measures could capture interactions between musical program material and sound reproduction systems that could explain in more detail the factors influencing preference choices. Of course, it is also of great interest to experimentally identify the most important auditory attributes that might predict the preferences for the sound stimuli presented in this study, but that issue is reserved for a subsequent paper. Nonetheless, this introduction begins with a brief discussion of these different components that play a role in the general assessment of spatial audio quality. These components are illustrated in the diagram 1, which shows the three domains between which relations are typically sought (as terms contained within the three boxes), and the quantitative data that are typically collected within each domain (italicized terms above each box). It seems most natural to suppose that physical phenomena give rise to sensory responses (hence the arrow connecting the boxes), which must intervene between the physical domain and the affective domain in which preferences are expressed are expressed as sentiments (as opposed to judgments, which may be regarded as correct or incorrect by the experimenter, as explained by Nunnally and Bernstein [69]). A more elaborate model could certainly be entertained (e.g., see [1]), which might assign evaluative weights to each of the assessed sensory attributes relevant to the formation of preferences, but a simplified overview will suffice for the present discussion.

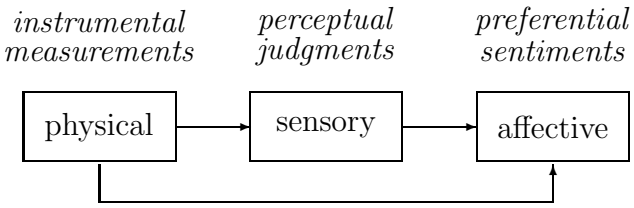


Diagram 1. Overview of the components that play a role in the general assessment of spatial audio quality (and many similar investigations).

There have been several good overviews of sound quality engineering of which the above offers only a pale reflection (see, e.g., [55] [27] [28]). But the above diagram makes a distinction that provides important introductory context for the current study, and the detail of particular interest is pictured in the diagram as the arrow passing under the box representing the sensory domain, and connecting physical and affective domains directly. The distinction is that the natural chain of phenomena, progressing from physical, through sensory, and finally to the affective domain, can be circumvented by attempting to develop relations directly between the physical domain (as quantified in instrumental measurements), and the affective domain (as quantified through preferential sentiments). Whereas it may ultimately make more sense to try to fit a complete model that relates observable data in the three domains, there are certainly reasons to try to relate instrumental measurements directly to preferential sentiments (cf. [38] [114]).

The first most important reason might be to avoid the expensive prospect of collecting many perceptual judgments that might not prove so useful in the overall scheme of the investigation. It might often be found to be better to manipulate stimuli to produce variation in terms of those physical measures that have been established as predictors of preference, and collect perceptual judgment data only for

a designed set of stimuli that could potentially be much reduced in size in comparison to the set of stimuli that might be selected without such guidance. Such an approach has been taken in studies of other sensory modalities as well, and particularly good examples may be found in food science, where it has been common to emphasize the practical importance of instrumental measurements (e.g., see [115], in which it was shown that with proper preprocessing, it is possible to map between preference and instrumental measurements made in physical units). This is precisely the motivation for the current study, which explores a set of physical measurements on the binaural signals associated with a small set of stimuli, for which preference choice data were already available.

Thus, the experimental context within which this study of physical predictors of preference is found can be summarized in terms of the two primary goals of the previously completed experiments: One goal was to determine whether different multichannel microphone techniques might be preferred for recording and reproduction of different types of musical performance. A second goal was to determine the experimental context within which such preferences might be observed, which it was hypothesized might depend upon the effects of changes in trial ordering. The novelty of the work presented in this paper is that it investigates whether physical predictors can be found to explain the observed dependence of listener preference on the type of musical performance being evaluated. Furthermore, there is a comparison between the goodness of fit achieved using the same predictors for the preference scale values observed for two independently tested groups of listeners who heard the same stimuli, but for whom the stimulus ordering differed substantially.

The way in which stimulus ordering differed between groups was designed to emphasize for one group stimulus comparisons of surround microphone techniques within blocks of trials which presented multiple versions of the same musical selection in trial after trial. Since the listeners in this group completed all trials for one musical selection before proceedings to blocks of trials for the next musical selection, this group was termed the *successive group*. Within the other group, listeners were presented with different musical selections on each trial, allowing them to focus not upon the particular differences between versions for a given musical selection, but rather allowing them to maintain a more global perspective on all the variations in spatial imagery they heard across trials. Since the listeners in this group heard different musical selections in a random order within each block of trials, this group was termed the *intermixed group*.

The way in which musical selection can influence pairwise preference choices differently for groups of listeners receiving different stimulus ordering has already been investigated in a previous paper [65]. The finding of a stimulus ordering effect in such preference testing is not without precedent. For example, Olive, et al. [116] obtained a similar result in a study of the influence of room acoustics on preferences for auditory imagery associated with a small set of loudspeakers. What is of interest here is whether an equation relating physical predictors to the previously derived preference scale values will be robust in the presence of such contextual dependencies. If most of the observed influence on preference choices can be explained in terms of differences between musical selections that exist in the binaural signals received by the listeners, then this may be taken as evidence that the prediction equation

developed here might generalize well across other methodoglogical variations. It will be shown via stepwise multiple regression analysis that the same predictors are found for both groups, and that there is no evidence that derived multiple regression coefficients differ between groups.

6.2 METHODS

6.2.1 Stimulus Preparation

Four solo piano pieces composed in the European concert musical tradition for this study: works by Bach, Schubert, Brahms, and a contemporary improvisation by Plaunt. The concept of genre or era in the history of Western composition is often hotly debated by musicologists, however, of interest in this study was not that a single composition might be taken to be symbolically representative of a specific era of composition, rather, of interest was that the musical selections represent a wide variety of the acoustical possibilities regularly encountered by audio engineers and producers recording "classical" piano music. In short, different musical selections activate the piano-hall acoustical system differently. As a result, it was hypothesized that some microphone techniques might be preferred for certain musical selections within the hall in question. Four surround microphone arrays were then selected: Fukada Tree, Polyhymnia Pentagon, OCT combined with a Hamasaki Square, and a SoundField microphone. All musical excerpts were performed in the same concert hall by a single musician, played on a single piano, and all four versions were captured simultaneously. The details of the recording procedure are well documented in [113].

6.2.2 Stimulus Presentation

Stimuli were presented through five full-range loudspeakers (Dynaudio model BM15A) in Multichannel Audio Research Laboratory (MARLab) at McGill University, with loudspeakers at a height of 1.2 m from the floor, and at a radius of 1.5 m from the listening position (slightly closer than that recommended in ITU-Recommendation BS.775-1 [4]. The ambient noise in the room was 27 dBA measured from the central listening position. Calibration of the loudspeaker levels was performed using a Brüel & Kjær type 2235 sound level meter with A-weighting and fast response. Each speaker output was individually calibrated to 78 dB SPL using a -18 dB FS pink noise input signal, which combined to give 85 dB SPL in total for all five speakers (that were selected from a larger sample of BM15A loudspeakers so as to match each other most closely in magnitude response). A Brüel & Kjær Head And Torso Simulator (HATS) was placed at the listening position and all 16 stimuli (4 musical selections and 4 microphone techniques) were recorded binaurally with the maniken's ears at a height of 1.2 m from the floor. These 16 binaural recordings were used for the physical signal analyses described in the next section of this paper.

6.2.3 Preference Choice Task

Each listening experiment was generally conducted in the following manner: Starting with an informal listening session, listeners were allowed to hear all four versions of the four piano performances, after which the experimental session was conducted, in which each listener completed four blocks of binary paired comparisons. Only two sound files were presented at a time, and subjects were asked to choose the preferred one. Listeners were instructed to select the file that they would rather

listen to for an extended period of time in their home over loudspeakers. During the experiments participants were not told which microphone arrays they were hearing, but were simply presented with two stimuli labeled A and B. Listeners were presented with each pair twice, but in opposite order, creating listening sessions comprising a total of 48 pairwise choices (four blocks of 12 preference choices). Both groups of listeners completed all four blocks of 12 preference choices, so that all heard all combinations of microphone techniques and musical selections; however, the order in which the trials were completed differed between the groups.

6.2.4 Successive versus Intermixed Trial Ordering

For the two groups of 18 musically experienced listeners the pairwise-comparison trials themselves were identical, as were the instructions that the subjects were given; however, for one group all trials for a given musical selection were completed in a single block, and then the experiment progressed to a block of trials for a different musical selection. This approach to trial ordering has been termed the *successive-treatment design* [103]. The second group of 18 listeners also completed four blocks of 12 preference-choice trials, but the musical selection was randomly assigned from trial to trial, so that the presentation of the four musical selections was distributed throughout the 48 trials. This approach to trial ordering has been termed the *intermixed-treatment design* [103]. Thus, for this group of 18 listeners, any effects due to sequential biases might be likely to be nullified, since the trial order was different for each listener. In contrast, the group of listeners receiving successive trial ordering had trial order randomized only within blocks of 12 trials, rather than over the entire 48 trials. Of course, in such a *successive-treatment design*, the order of

the blocks of single-musical-selection trials is a matter for concern. Therefore, the order in which the successive four blocks were completed was also randomized for the listeners in the successive-trial-ordering group.

6.3 ANALYSES AND RESULTS

6.3.1 Deriving Preference Scale Values

While direct scaling procedures are often used in audio-related research projects, binary paired comparisons were used here because, as explained in [63], indirect scaling using pairwise judgment does not rely on implicit and untested assumptions, and may be used to reveal relatively small differences existing among stimuli. In order to place a set of stimuli on a continuous psychological scale expressing the relative merit of each, the collected preference scores for the two groups of listeners were processed separately using a MATLAB function that estimates choice model parameters from paired-comparison data [64]. The derived preference scale values place each stimulus on an underlying continuous psychological dimension created specifically for the particular stimulus set being analyzed. Therefore, these values do not indicate the degree of preference among microphone techniques *per se*; rather, they express the relative preference for each stimulus within the entire set of 16 stimuli. The analysis of the obtained preference choice data, and the resulting scale values, are described more fully in [113].

6.3.2 Physical Measures

In an attempt to develop a quantitative model for predicting listener preferences from physical measures of the binaurally-recorded stimuli, 18 potential predictors were submitted to two separate stepwise multiple regression analyses, using as the

dependent variable the preference scale values obtained from the two groups of 18 listeners for the 16 analyzed stimuli.

The physical measures calculated were: ear signal incoherence (`ESI`), peak of signal envelope (`ENVMAX`), mean of signal envelope (`ENVMEAN`), standard deviation of signal envelope (`ENVSTD`), peak-to-mean signal envelope ratio (`ENVRATIO`), RMS (root-mean-square value) of mid signal (`RMSM`), RMS of side signal (`RMSS`), mid-to-side RMS ratio (`MSRATIO`), 1st–4th spectral moments of mid signal (`SMM1`–`SMM4`), 1st–4th spectral moments of side signal (`SMS1`–`SMS4`), ratio between high frequency and low frequency portion (with 250 Hz cutoff frequency) of mid signal RMS values, or mid-bass ratio (`MBR`), and finally, RMS ratio between high frequency and low frequency portion (with 250 Hz cutoff frequency) of side signal RMS values, or side-bass ratio (`SBR`). It should be noted that even though these last two variables might be thought to be related to the timbre of the stimuli, they also capture spatial information, since they are based upon the division of the binaural signals into M (mid) and S (side) components.

`ENVRATIO` was included to in attempt to capture the dynamic variation presented in a musical piece. The idea for this measure was based upon the observation that some music program had more dynamic variation (viz., level difference between quiet and loud parts) which, combined with certain surround microphone technique, affected the timbral and spatial impressions formed by listeners. `MSRATIO` was included to be a simpler substitute to `ESI` or IACC. Similar measure to this is *lateral energy fraction* proposed by Barron and Marshall [20] to predict auditory source width, which is calculated as an energy ratio of impulse response recorded with bi-

and omni-directional microphones with different time frames for each microphone. Unlike lateral fraction, MSRATIO does not take the time frame portion into account. SBR and MBR were included based on the comments from listeners that some musical selections sounded too muddy in the low frequency region. The cutoff frequency of 250 Hz was chosen by authors based on listening skills derived from Tonmeister training. Later, this choice was justified visually on the basis that the spectrograms of the stimuli showed the difference between the musical programs quite clearly.

6.3.3 Multiple Regression Analysis

In a pilot analysis, only the first 16 physical measurements listed above were used. For the intermixed group, the stepwise regression analysis began with a model that included only ESI as the predictor, with $R^2 = 0.461$, and no significant improvement was seen with further inclusion of other terms. On the other hand, for the successive group, the stepwise analysis began with a model equation using only MSRATIO as the sole predictor, with $R^2 = 0.662$. Again, no substantial improvement in the fit was obtained by adding any other term to the prediction equation, but their was some evidence that this analysis was ill-conditioned, most noteably due to the high correlation between ESI and MSRATIO (which was $r = 0.955$). Therefore, the MSRATIO was excluded to avoid collinearity (see Draper and Smith [117] for an explanation of this difficulty in multiple regression analysis). A subsequent pilot analysis showed that ESI values together with RMSS values also predict preference scale values well for the successive group ($R^2 = 0.759$) in the absence of the MSRATIO term. But the RMSS values glossed over a potentially important frequency dependence that might allow differences between different program materials to be

introduced into the prediction equation more strongly. Hence, after these pilot analyses, it was deemed potentially useful to add two additional independent variables, `MBR` (mid-bass ratio) and `SBR` (side-bass ratio), In that these are RMS ratios between high frequency and low frequency portion (with 250 Hz cutoff frequency) of mid and side signals, they were thought to potentially encode the interaction between program and microphone technique, which was showing up as substantially large in the regression residuals. The regression residuals also indicated that a curvlinear relation was not being captured in the linear regression equations of the pilot analyses, and it was decided to run additional analyses on the log-transformed preference scores as dependent variables, termed hereafter `LOGP1` and `LOGP2`.

Although, `MSRATIO` could have predicted preference reasonably well in the absence of the `ESI` term (with $R^2 = 0.447$ for the intermixed group and $R^2 = 0.662$ for the successive group), adding two new variables and switching to a log-scaled dependent variable for the stepwise regression analysis gave a model equation using the same two parameters for both groups of listeners: `ESI` and `SBR`. The coefficient of determination for the successive group was $R^2 = 0.858$ and that for the intermixed group was $R^2 = 0.828$. In figure 6–1, the obtained preference scores are plotted on the predicted preference scores for both groups, and the prediction equations were formed as follows (labelled as equation 6.1).

$$\log_{10}(P_1) = 2.03 \cdot E - 0.21 \cdot S + 1.37$$

$$\log_{10}(P_2) = 2.33 \cdot E - 0.22 \cdot S + 1.21$$

$$(E = \texttt{ESI} \;\; and \;\; S = \texttt{SBR}) \tag{6.1}$$

6.4 DISCUSSION

For both groups of 18 subjects, the simple regression equation relating 16 stimulus ESI values to the log of the average preference scale values accounted for more than 70% of the variance. But adding SBR to the prediction equation increased the proportion of variance for which the model accounted from .716 to .858 in the case of the successive group, and from .716 to .828 for the intermixed group. In both cases the improvement in fit was significant at $p < .05$. This is despite the fact that the correlation between the dependent variable and `SBR` was not significant for either group. This detail is worth discussing here, since it suggests that `SBR` may be operating as a *suppressor variable* within the prediction equation (see [80, Chap. 5] for an explanation of this phenomenon).

That is to say, that even though `SBR` by itself cannot predicting the outcome, it improves the power of `ESI` to predict the outcome, and achieves a better fit by "cleaning up" the differences between musical programs that may be not reflected in `ESI` values. Another way to express this would be that `SBR` modulates how variation in `ESI` is mapped to changes in preference for microphone technique due to the content of the program under evaluation. This is of course only a speculation, as there is no proof here for such a causal relation in the current study. However, the speculation makes sense, especially when partial correlation values are considered[1] .

[1] Partial correlation is a conditional relationship between three variables, X, Y, and Z, that is measured as ρ, the portion of the relationship between Y and Z that

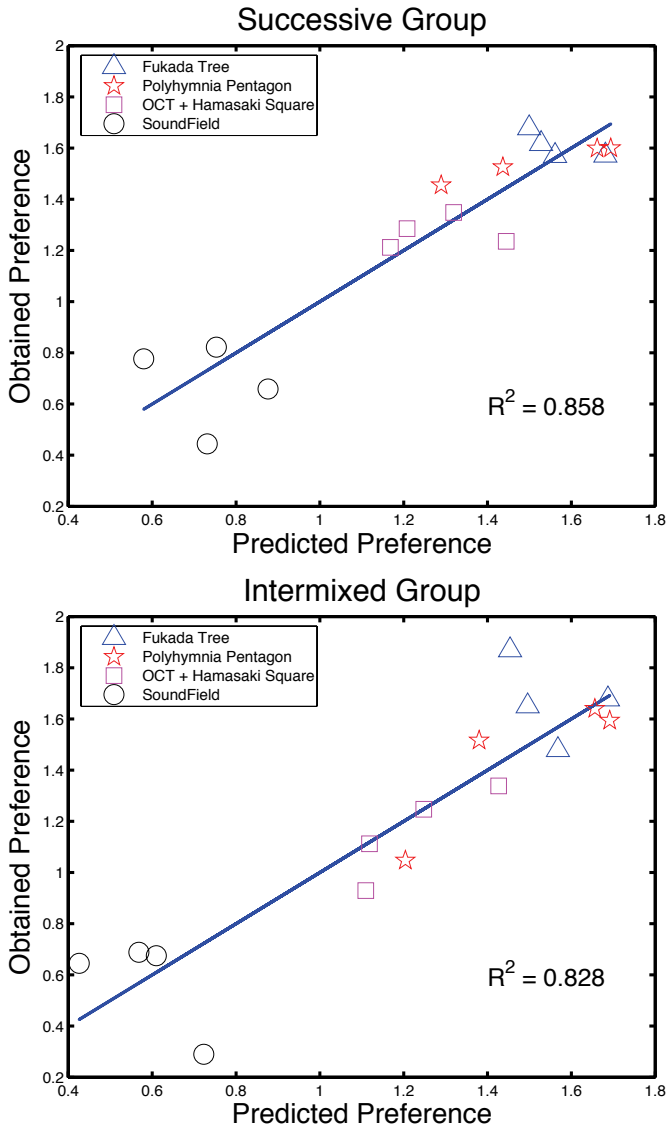


Figure 6–1: Results of multiple regression analyses run sepatarely on two independent groups of 18 subjects, one receiving trials according to the *successive-treatment design*, the other receiving trials according to the *intermixed-treatment design*. Obtained preference (log of the average preference scale values) is plotted on the predicted preference values based upon a two-term regression equation that included ESI and SBR values for each of the 16 stimuli.

In particular, it is worth noting that the Pearson correlation between ESI and the outcome (obtained preference) is high to begin with (at $r = .85$ for both groups), but this relationship is strengthened after removing the portion of the relationship between these two that has no linear relationship to SBR. The remaining variance in ESI shows a stronger relationship with the outcome for both groups of listeners after accounting for SBR, an increase which is measured as $\rho = .92$ for the successive group, and as $\rho = .90$ for the intermixed group. Also, in the case of the successive group, the Pearson correlation between SBR and the outcome is only $r = .20$, but taking into account the relation between ESI and the outcome, the partial correlation between SBR and the outcome becomes strongly negatively at $\rho = -.71$. For the intermixed group, the Pearson correlation between SBR and the outcome is only $r = .23$, but taking the effects of ESI on the outcome into account, the partial correlation between SBR and the outcome becomes strongly negatively at $\rho = -.62$.

The reason for focussing upon these partial correlation values is that the current analysis is only one component of an exploratory investigation of preferences for multichannel loudspeaker-reproduced auditory imagery, in which identification of the role played by physical predictors is quite important for potential practical applications (as also seen in the extensive exploratory study of physical predictors of loudspeaker preferences undertaken in [36, 37]).

---

has no linear relationship to X. See [80, Chap. 7] for an explanation of this statistical measure.

Also worth discussing here is the potential relation between variation in the most predictive physical measures and variation in their associated auditory attributes. While it is best to base such a discussion upon empirical data, collection of which is already under way, it is also reasonable to suppose that ESI would be related to the attribute termed auditory source width (ASW). The auditory attribute modulated by variation in SBR remains to be determined, but it is hoped that ongoing investigation will reveal the perceptual means by which this parameter is contributing to the interaction between musical program and prefered surround microphone techniques.

6.5 CONCLUSION

Various physical measurements were made on the binaurally recorded signals associated with multichannel loudspeaker reproduction of a set of four piano performances. Those measurements were analyzed through a stepwise regression in order to generate a prediction model for previously obtained preference scale values derived for the same stimuli. Two measures, ESI and SBR, were chosen as two best predictors for log values of obtained preference scores, with R^2 values higher than 0.8 for both of two independently tested groups of listeners. It was inferred from the results that while ESI accounted well for the variance in preferences due to surround microphone techniques, SBR itself worked as a suppressor variable that helped to make the prediction equation more sensitive to differences between musical program in their interaction with the four tested microphone techniques. This RMS ratio between low frequency and high frequency portions of the side signals derived from binaural signals, SBR, added a frequency-dependent component to the prediction equation, which alone did not predict preference well, but in combination with ESI produced

the best fitting result for both of the groups tested. Finally, the regression equations derived for the two groups were practically identical, despite the differences that exited in methods of data collection used for the two independently tested groups of listeners.

CHAPTER 7
Deriving Physical Predictors for Auditory Attribute Ratings Made in Response to Multichannel Music Reproductions

This chapter contains the exact the copy of the paper published in the proceedings of the 123^{rd} International AES convention held at New York, USA in October 2007, which was written in collaboration with William L. Martens.

ABSTRACT

A group of 8 students engaged in a Tonmeister training program were presented with multichannel loudspeaker reproductions of a set of solo piano performances, and were asked to complete two attribute rating sessions that were well separated in time. Five of the 8 listeners produced highly consistent ratings after a 6 month period during which they received further Tonmeister training. Physical predictors for the obtained attribute ratings were developed from the analysis of binaural recordings of the piano reproductions in order to support comparison between these stimuli and other stimuli, and thereby to establish a basis for independent variation in the attributes to serve both creative artistic goals and further scientific exploration of such multichannel music reproductions.

7.1 INTRODUCTION

It is a fundamental assumption underlying modern techniques for perceptual evaluation of musical sound reproduction that listeners are able to analyze their complex auditory percepts in terms of separable attributes. Furthermore, this assumption rests upon a number of suppositions that provide a foundation for experimental work in this area. One supposition is that these attributes are relatively permanent perceptual characteristics of reproduced musical sound that will remain reliable over time for a given stimulus domain. Another supposition is that these attributes are grounded in physical characteristics of reproduced musical sound that may be measured for a set of stimuli, and these measures should discriminate between stimuli in a manner that parallels how ratings on the attributes discriminate between the percepts associated with those stimuli.

Although there are other aspects besides these two suppositions that are also quite important in the process of attribute identification (see, e.g., [33]), these two are regarded as quite essential to experimental work, especially since both suppositions are related to hypothesis testing, as will be discussed in the subsection of this introduction entitled "Two suppositions." The quantification of auditory impression via attribute ratings is influenced by many variables, some of which can be experimentally controlled [28], and others that must be treated as unavoidable contextual factors [112]. In the study to be described in this paper, the experimental variable that was under direct control was the multichannel microphone technique that was used to record a selection of solo piano performances. Another important factor here was the selection of musical program material to be used in evaluating

the results of using the microphone techniques to be evaluated. As previous reports on this project have already given more in-depth introduction to these issues (e.g., [86] [118]), this introduction includes only a brief presentation of the key questions on which this paper will focus. The questions to be addressed relate to the two suppositions described in the following subsection.

7.1.1 Two suppositions

The supposition that subjective ratings made on a given set of attributes can provide an unchanging description over time for the perceptual characteristics of reproduced musical sound lends itself to direct experimental test. The related hypothesis can be stated simply in terms of reliability, without any need to address the validity of the attributes in question, as follows: The ratings that a given listener produces on one occasion for a restricted set of stimuli will be highly correlated with the ratings that same listener produces on another occasion, removed in time from the first so as to represent an independent assessment of the characteristics of those stimuli. Confirmation of this correlation for each listener tested will provide a stronger test of the hypothesis than does the single correlation that can be measured for the combined ratings of a group of listeners on two separate occasions. The latter "group" test also does not allow any determination of which listeners are producing reliable ratings, and which exhibit inconsistent ratings over time.

The supposition that each of a given set of attributes is grounded in some measureable physical characteristic of reproduced musical sound also lends itself to direct experimental test. Here, though, it is the validity of the attributes that is in question, and the related hypothesis can be stated as follows: Physical measures made

on each item in a restricted set of stimuli predicts variation in the ratings made on the proposed perceptually most salient attributes that purportedly characterize the complex auditory responses associated with those stimuli. Without the benefit of such physical measures on experimental stimuli, it is difficult to be sure why certain sets of attribute ratings are correlated with each other. In effect, the issue here is to determine whether attribute ratings that are correlated for a given set of stimuli are correlated because of coincidental relations within the restricted set of stimuli, or whether those attribute ratings are inherently correlated because they represent only slightly contrasting verbal perspectives on a single underlying auditory attribute. Ultimately, having experimentally verified physical predictors for the various attributes thought to be salient for a given set of stimuli has many benefits, not the least of which is how they may aid in stimulus selection in subsequent experiments. For example, causal relations between predictors and percepts can be examined through factorial combinations of predictor values that defeat inter-stimulus correlations. Furthermore, having a number of experimentally verified physical predictors can serve to clarify for listeners what attributes they will be asked to subjectively rate, since there may be a need to provide examples of stimuli at each extreme of the attribute scales the listeners will be using.

7.2 METHODS

7.2.1 Listeners

A total of eight masters students in the Sound Recording program of McGill University participated in the listening experiments. While these students could not be regarded as experts either in sensory evaluation nor in sound recording practice,

they were all engaged in the training that follows the *Tonmeister* tradition, in which they develop skills in microphone placement and in aural evaluation of the results. Furthermore, they were all engaged in regular sessions of timbral ear training [119], and may be regarded as having acquired special abilities to make discriminations and distinctions between reproduced sound stimuli. Therefore, they certainly could not be characterized as naive listeners. However, they could neither be characterized as experienced assessors with regard to such perceptual tasks as were required for the current study. Suffice it to say that they were motivated to do well on these tasks that appeared to be related to the skill sets that they desired to develop through their studies, and this position was clearly expressed within group discussions during debriefing.

Although results obtained from such trained listeners may not be consistent with results obtained using untrained listeners typically reported in the literature on perceptual audio evaluation [120], the elicitation and ratings made by *Tonmeister*-trained listeners may provide a more comprehensive set of auditory attributes. *Tonmeister*-trained listeners have what might be termed the "*Tonmeister* bias," which has sensitized them to spatial attribute differences presented via multichannel sound reproduction for which naiive listeners have fewer clear distinctions. Suffice it to say that it was their refined assessment of the perceptual consequences of using different microphone techniques that was of primary interest in the current program of experimental study.

7.2.2 Direct ratings on selected attributes

Five attribute scales were previously constructed on the basis of the results of a verbal elicitation task undertaken by the above-described eight listeners [86]. The attribute scales were anchored by five pairs of bipolar adjectives upon which the eight listeners reached some consensus, and included the following adjective pairs:

- Wide ↔ Narrow
- Focused ↔ Diffused
- Tight-Bass ↔ Muddy-Bass
- Sharp ↔ Dull
- Distant ↔ Close

The same eight listeners made direct ratings of perceived magnitude associated with each attribute for each of 32 relatively short stimuli. Analysis on those ratings showed that two principal components accounted for most variance of five attributes [86]. The first principal component was associated with attribute **WIDTH** while the second was with both attributes **BASS-TIGHTNESS** and **SHARPNESS**. Since **BASS-TIGHTNESS** and **SHARPNESS** were relatively highly correlated and appeared as similar in the principal component space, **BASS-TIGHTNESS** was taken as representative of the attribute associated with the second principal component (especially since x was regarded as likely representing a timbral distinction rather than the spatial distinctions upon which listeners had been instructed to focus).

Six months after initial direct ratings on those five attributes were completed, the same eight listeners were invited to rate the same stimuli on a slightly revised

set of five attribute scales: Three of these were the same as in the first session, and included **WIDTH**, **FOCUS**, and **BASS-TIGHTNESS**. Two new attributes were included in the subsequent rating sessions, and these were **BRIGHTNESS** and **LISTENER ENVELOPMENT** (LEV). The anchor "Sharp" was replaced by "Bright" because of an ambiguity that became apparent after the completion of the first session. The term "Sharp" had been used by listeners as the antonym of "Dull" within the context of timbral balance; however, "Sharp" also could be related to a spatial impression regarding the degree to which an auditory image is defined and not blurred in its spatial position. Thus the attribute **SHARPNESS** was replaced by **BRIGHTNESS** which was anchored by the bipolar adjectives, "Bright $\leftrightarrow$ Dark." It is worth noting, however, that since the definition of the attribute scale for **SHARPNESS** was defined for the listeners as **the variation in tone coloration associated with an increase in high-frequency content relative to low-frequency content in the sound source (piano)**, the risk of such confusion should have been minimal. Nevertheless, it was of interest to compare ratings from trained listener ratings on **SHARPNESS** and **BRIGHTNESS**, especially when an identical definition was given to the listeners before each listening session, and when a 6 month interval has elapsed between the two listening sessions, so that the replication of the written definition between the two terms was not obvious to the listeners.

The **LISTENER ENVELOPMENT** (LEV) scale then effectively replaced the one remaining previously elicited attribute scale of **DISTANCE** that had appeared in the selected group of five attributes in the first elicitation results [86]. Indeed,

listeners selected bipolar adjective pairs related to LEV slightly less frequently than **DISTANCE**. But, since the attribute **DISTANCE** was quite highly correlated with **WIDTH**, it came to authors' attention that LEV ratings might be used to differentiate between spatial impressions associated with the sound source rather than the environment in which that source was located. This distinction is often made in multichannel music reproduction, but is more difficult to make in the case of piano music, because this musical instrument as a sound source is more spatially extended than most. Nonetheless, this term was employed in the second rating session since in debriefing some listeners indicated that LEV was varying within the presented set of stimuli, though the authors' initial evaluation indicated that LEV was not varying greatly for the set of stimuli.

As in the first session, listeners also gave ratings on their relative preference for each of the 8 piano performance excerpts that were presented. Thus for the second round of listening sessions, each listener gave ratings for 32 stimuli on this one sentiment, in addition to the five attribute scales anchored using the following adjectives:

- Wide ↔ Narrow
- Focused ↔ Diffused
- Tight-Bass ↔ Muddy-Bass
- Bright ↔ Dark
- Enveloping ↔ Non-Enveloping

7.3 RESULTS

7.3.1 Comparing attribute ratings over time

Pearson correlation coefficients between the first and second ratings were calculated for each subject and each attribute. The observed correlations are presented in Table 7–1, with **bold font** representing the correlation coefficients that were lower than the criterion for statistical significance (at probability $\alpha < .05$ of incorrectly retaining the null hypothesis in each case). Such small correlation coefficients might indicate either that listeners in these cases were inconsistent in how they understood the attributes on which they were required to make their ratings, or that they simply were not able to make consistent magnitude estimates for some attributes, though they might have understood well the meaning of the anchors defining the extremes for each attribute scale. Regardless of the reason, three subjects, S1, S3 and S5, were separated out from the other five as relatively poor in producing ratings that matched their previous ratings. Since combining such inconsistent perceptual responses provides a poor definition of the responses to be related to associated physical measures, the ratings from these three relatively inconsistent subjects were excluded from the subsequent regression analysis designed to examine those potentially predictive physical measures. It is worth noting that 6 out of 8 listeners produced significantly correlated ratings for the **SHARPNESS** and **BRIGHTNESS** scales even though the bipolar adjectives given to listeners as anchors were different. For ratings on **DISTANCE** and LEV, the five selected listeners showed no significant correlation.

Subject	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**
Preference	**.066**	.377	**-.083**	.500	**.248**	.554	.834	.673
Width	.495	.674	.382	.526	**.029**	.576	.452	.542
Focus	**-.138**	.375	**.165**	.488	.379	.583	.666	.574
Bass Tightness	**.029**	**.243**	.550	**.139**	**.315**	.542	**.137**	.716
Sharpness & Brightness	.531	**.244**	.544	.871	**.287**	.535	.511	.537
Distance & LEV	**.117**	**-.089**	**.315**	.478	**.120**	**.141**	.737	.435

Table 7–1: Pearson correlation coefficient matrix between first session ratings and second session ratings. Bold font is used to show which correlations were not significant (less than the the critical value $r = 0.345$, for a two-tailed t-test, and with $df = 30$ at $alpha = .05$). The first four rows show correlation coefficients between the two sessions of ratings on identical attributes. The fifth row contains correlation coefficients between Sharpness ratings and Brightness ratings while the sixth row contains those between Distance ratings and LEV ratings.

7.3.2 Physical predictors for attribute ratings

It was shown in a previous study [118] that a quantitative model could be developed that relates listener preferences for a related set of stimuli to two physical measures that were termed Ear Signal Incoherence (`ESI`) and side-bass ratio (`SBR`). The observed coefficient of determination in predicting the obtained log preference values calculated for two independently tested groups of 18 listeners was $R^2 = 0.858$ and $R^2 = 0.828$. Later a pilot study employing experimenter-selected attributes was completed in which the same stimuli were presented [121]. The results showed that the physical measure termed `ESI` was strongly correlated with ratings of the perceived width of the reproduced sound image ($r = 0.77$), which was expected since this broad-band interaural cross-correlation measure has often been shown to be highly predictive of variation in this attribute. It was also found that ratings on the tightness or muddiness of the bass imagery covaried (at $r = 0.83$) with the

values that the stimuli took on a physical predictor termed "Side-Bass Dominance," or SBD. The predictors derived from these prior results aided in stimulus selection for the current study, which aimed to present a more balanced distribution of these two physical predictors that had been presented initially.

So, in designing the stimulus set for the current study, it was desired that the the stimuli should exhibit values on these two physical measures (ESI and SBD) that showed low correlation across the entire stimulus set. This is in contrast to the stimulus set used in the previous study, a set within which stimuli with high ESI values also tended to have high SBD values. In order to test the independent power of the two predictors, two short excerpts taken from each piano performance were selected so that one would have a relatively high SBD value, and the other would have a relatively low SBD value. In the previous study, only one excerpt of each of the four piano performances was presented for each of the four microphone techniques, making a total of 16 stimuli that exhibited a relatively high correlation between their values on these two physical measures. Instead, the set of 32 stimuli employed in the current study exhibited a relatively lower correlation between ESI and SBD.

Foundations for new predictors

There was a need to re-examine the physical measures used in prediction of preference ratings developed in the previous study [118], since the two supposedly associated attributes did not show such a strong correlation with the two physical measures, ESI and SBD. Of course, the attribute ratings that were subsequently collected here were made in response to a new set of stimuli, and with a new group of listeners. There were several important reasons to develop new measures, one of

which was that there previously was not enough data collected on a well-established set of attributes. Also, the durations of the newly selected stimuli were shorter than those of the previous stimuli, which had exhibited some considerable variation in the measured values over time due to the natural progression of the piano performance. It was hoped that by using shorter excerpts here, there would be less temporal variation in the measures of the stimuli, and a better chance to observe a close fit between newly-developed predictors and the newly-collected attribute ratings. For these new, perhaps more tractable stimuli, a new set of physical predictors was sought, along with a new foundation for these predictors.

In previous studies [2][114][122][74][48][76], substantial amounts of effort have been devoted to the endeavor to relate subjective preferences and/or individual attribute ratings to physical measures of reproduced musical sound stimuli. The results of these previous studies provided a strong foundation on which following relevant research could stand. For example, a well-accepted prediction model for perceived loudness [48] and an elaboration on the standardize predictor for sharpness [2] have been adapted in the current study in an effort to measure the stimulus magnitude of auditory attribute scales (in order to replace the potentially more expensive and relatively less consistent human responses). Such prediction models were commonly based on two psychoacoustic phenomena; frequency selectivity and masking within human audition. Such phenomena have been shown to have strong basis in Critical Band function. Subsequently, the current model for predicting obtained ratings of attribute **WIDTH** and **BASS-TIGHTNESS** were also based on this

psychoacousticaly-informed approach, which utilizes modelled Critical Bands, combined with application of those bands that has been vaildated by the results of many previous studies (see [123] for review).

Predicting WIDTH

Binaurally recorded signals were split into 25 Equivalent Rectangular Bands (ERBs) using Auditory Toolbox 2, a Matlab toolbox developed by Malcolm Slaney[124]. The center frequencies of each ERB correspond to the standard ISO values: 63, 80, 100, 125, 160, 200, 250, 315, 400, 500, 630, 800, 1000, 1250, 1600, 2000, 2500, 3150, 4000, 5000, 6300, 8000, 10000, 12500, and 16000 Hz. The correlation between left-right ear signals within each band was calculated based on the equation 7.1. SLi and SRi represent the Standardized Left and Right ear signals at ith ERB respectively. These correlations measured on each ERB were summed to a single number to represent a predicted magnitude of **WIDTH** for a given duration of each stimulus (w in the equation 7.2). This prediction model did take into account any additional weightings associated with each frequency band, though it might be possible to obtain a better-fitting prediction equation by applying a proper weighting function as suggested by Morimoto [125] and Mason [74]. In a pilot study, a slight boost between 100Hz and 1000Hz gave a slight increase in the correlation coefficient. However it was not a significant increase and the two frequencies were derived by brute-force iterations, rather than any psychoacoustic data. Therefore the validation and implementation of a frequency weighting function for the prediction of **WIDTH** will be left for a subsequent experiment.

$$f(i) = \sum_{k=1}^{N} \frac{SLi(k) \cdot SRi(k)}{N-1} \tag{7.1}$$

$$w = \sum_{i=1}^{25} f(i) \tag{7.2}$$

Predicting BASS-TIGHTNESS

The prediction model for **BASS-TIGHTNESS** was implement in same way as that for **WIDTH**, except that the ear signal correlations up to only the 9th EBR (with center frequency of 400Hz) were summed to generate a simple metric for the predicted magnitude of **BASS-TIGHTNESS** (Bt in equation 7.3) Even though this prediction model was similar to the prediction model of **WIDTH**, predicted values for **BASS-TIGHTNESS** were not so highly correlated with ratings of **WIDTH** ($r = 0.4167$). Usually the level of signal has been known to modulate perceived magnitude of spatial attributes, especially at low frequency [125]. Probably that was why the previously-developed physical predictor, termed `SBD` in [118], could be used to **BASS-TIGHTNESS** well for the longer stimuli that were presented in that previous study.

$$Bt = \sum_{i=1}^{9} f(i) \tag{7.3}$$

The upper panels in Figure 7–1 show the results of the derived prediction model for both **WIDTH** and **BASS-TIGHTNESS**. The abscissa of each plot in the

upper panels of Figure 7–1 represents the averaged mean rating magnitudes for the two predicted attributes, **WIDTH** (on the left) and **BASS-TIGHTNESS** (on the right). The ordinates represent the predicted magnitude based upon the regression results for the two attributes. The mean obtained magnitudes for these attributes were based upon 10 ratings (two trials each for each of five listeners) on each attribute. The correlation coefficient between obtained mean ratings and predicted **WIDTH** was 0.873 and the correlation coefficient between obtained mean ratings and predicted **BASS-TIGHTNESS** was 0.787.

7.3.3 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) was employed to examine the relationship between four of the sets of collected attribute ratings (**WIDTH**, **FOCUS**, **SHARPNESS** and **BASS-TIGHTNESS**). The first two components accounted for most of the variance (93%) in the four attribute rating datasets. The submitted ratings included results from the two separated listening sessions made by the five relatively consistent listeners (each listener making 10 ratings on each attribute). Similar to the previous study's result, principal component 1 (PC1) was strongly associated with two attributes, **SHARPNESS** and **BASS-TIGHTNESS**, while the second compoent (PC2) was associated primarily with the **WIDTH** attribute. The correlation coefficient between PC1 and **SHARPNESS** was $r = 0.926$ and the correlation coefficient between PC2 and **WIDTH** was $r = 0.989$.

The two bottom panels in Figure 7–1 show the relationships between predicted rating magnitudes for two attributes and the two sets of component scores from the PCA run on four salient attributes for the entire set of 32 stimuli. The correlation

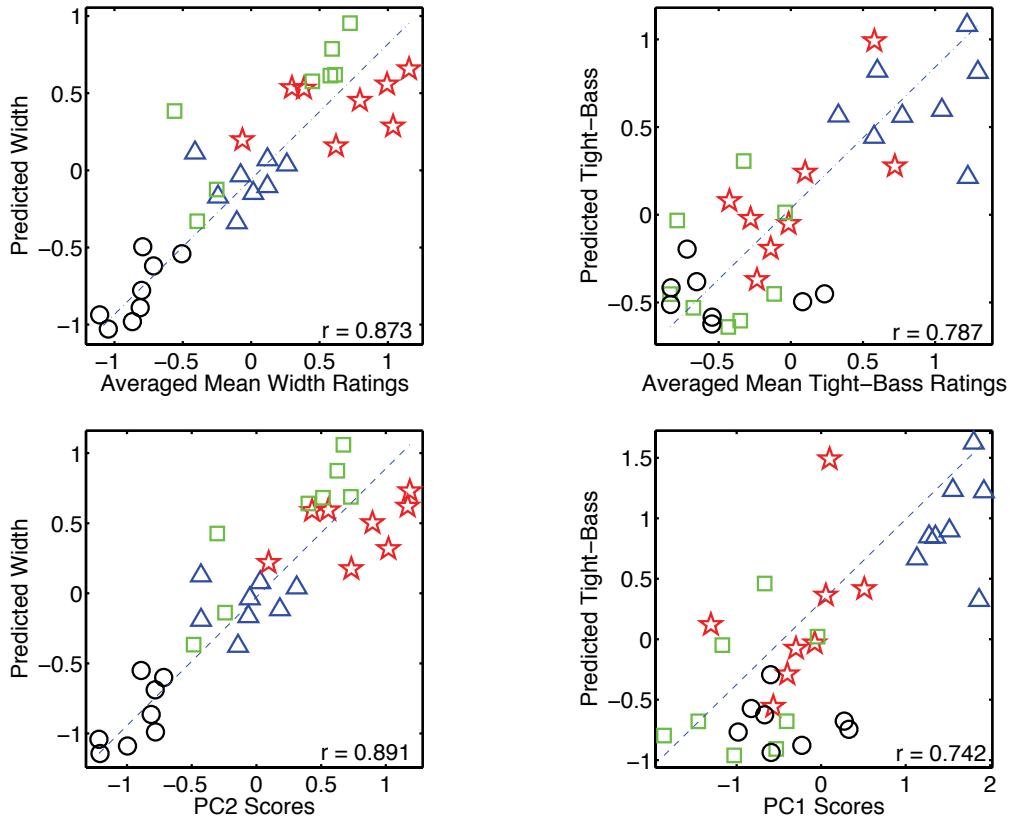


Figure 7–1: **[Upper left panel]** Scatterplot of predicted magnitudes of **WIDTH** calculated via equation 7.2 vs. mean ratings of **WIDTH** for each of 32 stimuli, and each of 5 listeners, in each of 2 sessions (hence the "Averaged Mean" axis labels). The plotting symbols used here made no distinction between listeners or the 8 musical programs to which they listened; rather the symbol shape codes only which microphone technique was employed for each rated stimulus: blue triangle for Fukada Tree, red pentacle for Polyhymnia Pentagon, green square for Optimized Cardioid Triangle with Hamasaki Square, and black circle for SoundField MKV. **[Upper right panel]** Scatterplot of predicted magnitudes **BASS-TIGHTNESS** calculated via equation 7.3 vs. mean ratings of **BASS-TIGHTNESS**, again for 32 stimuli. **[Lower left panel]** Scatterplot of predicted magnitude of **WIDTH** vs. the second principal scores of 32 stimuli derived from four salient attributes. **[Lower right panel]** Scatterplot of predicted magnitudes of **BASS-TIGHTNESS** vs. the first principal scores of 32 stimuli.

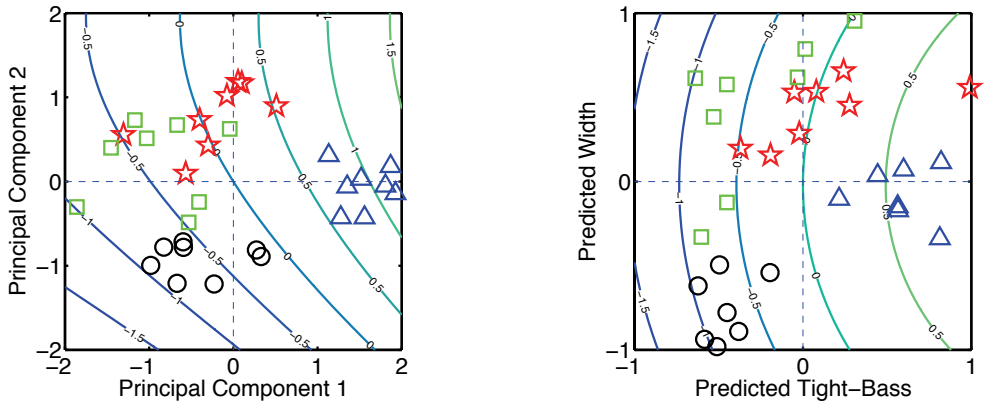


Figure 7–2: **[Left panel]** Scatterplot of PC1 vs. PC2 with its derived iso-preference contour showing the relationship between scores calculated for 32 stimuli and the mean preference ratings for those stimuli. **[Right panel]** Scatterplot of **Predicted Width** vs. **Predicted Bass Tightness** for 32 stimuli with the associated iso-preference contour.

coefficient between PC1 and predicted **BASS-TIGHTNESS** was $r = 0.742$ and the correlation coefficient between PC2 and predicted **WIDTH** was $r = 0.891$. These results show that scores on the first principal component (PC1) covary with predicted **BASS-TIGHTNESS**, while the scores on PC2 have a stronger relation with the predicted **WIDTH**.

7.3.4 Principal components and preferences

Subsequently, an attempt was made to predict preference ratings from stimulus values on the two sets of principal component scores via stepwise regression using a quadratic response surface model. In this stepwise regression, quadratic terms for each of the independent variables, PC1 and PC2, were added to investigate the gradient of a preference surface in the principal component space. This relationship between preference and the scores on the first two principal components can

be visualised via the iso-preference contours shown in the left panel of Figure 7–2. The first principal scores contain the dominant portion of variance explaining preference ratings and the second principal component scores contribute some significant modulation to the preference response surface. The proportion of the variance in preference ratings that was accounted for by the stepwise regression was $R^2 = 0.77$

7.3.5 Attribute predictors and preferences

The right panel of the Figure 7–2 shows the derived preference contour in the space defined by the average values predicted for each of 32 stimuli on each of the two sets of attribute ratings that were associated with the principal axes observed in the PCA result. Again, quadratic terms for both predictions were used in a stepwise regression. Unlike the previous iso-preference contour fit to PC1 and PC2, which both significantly contributed to the successful prediction of preference, only predicted magnitudes of **BASS-TIGHTNESS** account for the bulk of the variance in preference. Including predicted **WIDTH** in the model did not significantly increase the total amount of variance for which the model could account (with a coefficient of determination $R^2 = 0.71$).

Although **WIDTH** was clearly well described by the physical measure designed to predict ratings on this attribute, the predicted **WIDTH** values did not significantly modulate preference. This finding is consistent with recently reported results showing that the attribute ratings themselves were not strong predictors of preference for the same stimuli [86]. It was hypothesized that the failure to show the relation between attribute ratings on **WIDTH** might have included too much magnitude estimation error on the part of the listeners, and that the physically predicted values

associated with the **WIDTH** attribute would provide a better fit to the preference ratings. This hypothesis was not supported here. Therefore, this highly salient attribute describing the auditory character of the current stimulus set is probably not playing an important role in how listeners form their preferences here. On the other hand, there is strong support for the hypothesis that **BASS-TIGHTNESS** had a significant influence on the formation of listener preferences for the current set of 32 stimuli.

7.4 CONCLUSION

It was hypothesized that the attributes found most salient in previous studies could be reliably rated in well separated sessions by 5 out of the 8 listeners tested here. That is to say that ratings produced by a listener on one occasion for a restricted set of stimuli were found to be correlated significantly with the ratings that the same listener produced on another occasion 6 months later period. This finding supports the supposition that subjective ratings made on a given set of attributes can provide an unchanging description over time for the perceptual characteristics of reproduced musical sound, at least by a majority of listeners.

It was also hypothesized that each of a given set of attributes could be validly associated with some measureable physical characteristic of the reproduced musical sounds for which reliable ratings on those attributes were collected. Two physical measures made on each item in the restricted set of stimuli presented here were found to predict variation in obtained ratings. These ratings were made on two of the most salient attributes that were thought to characterize the complex auditory responses associated with those stimuli; however, ratings on only one of the two attributes,

BASS-TIGHTNESS, was found to be a significant predictor of preference ratings for the same stimuli. When the other attribute, **WIDTH**, was added to the prediction equation for preference, there was no significant improvement in the proportion of variance for which the equation could account. Nonetheless, having a number of experimentally verified physical predictors will certainly aid in stimulus selection in subsequent experiments using related stimuli.

CHAPTER 8
The effect of context on sound quality evaluation

8.1 The wherefores

A big assumption in any perceptual evaluation is that a total impression is the sum of separable sound characters of which the listeners can perceive and express the relative strength. Validity of this assumption would be fulfilled if a listener's affective or perceptual response is not altered by non-experimental variables such as previous experience, current emotion, effects from the previously presented stimulus, noises in the signal chain and the acoustical environments, etc. In general, these variables have been treated as "nuisance variable" which, nevertheless, would not influence the listener response regardless of their presence. In contrast, there have been several claims that such nuisance or error variables might cause an effect that is not small enough to be neglected as random variables [111], which makes it difficult to measure the consistent perceptual or affective response of the listener and eventually prevents the experimenter from deriving a reliable conclusion. These nuisance variables are often referred to as **biases**. The modern literature of the perceptual evaluation, thus, endeavors to identify and analyze the effect of such variables, or biases, on the main experimental variables. Poulton [126] investigated various kinds of biases and divided them into three categories as summarized in [28, Chap. 4.2.4]:

- Contraction bias caused by the subject's tendency to be conservative so that large differences are underestimated and small differences are overestimated
- Bias caused by (un)familiarity with units of magnitude
- Bias caused by unfamiliarity with the mapping of the responses to the stimuli

Among many biases, a particular effect caused by the given context became attentive in the reproduced sound quality evaluation. It is probably because any listening experiment involves a relatively long period during which the subjects are exposed to a sound field; it eventually makes it easy for the subjects to adapt to the environment where they are. For example, when a listener compares two sound fields, the first sound field might modulate the internal standard of the subject; the second one will be judged based on that adapted standard. Therefore, it is of importance to conduct a sound quality evaluation considering the contextual influence on the listeners' response. Rumsey [110] asserted that "if we are to stand a chance of being able to predict factors such as listener preference or liking on the basis of expert ratings of descriptive quality attributes, then a reliable means of accounting for the context dependencies of such matters needs to be devised." This dissertation research, consequently, has experimentally scrutinized the "nuisance variables" and found that two contextual variables had influence upon the main research results, which will be summarized and introduced in the following subchapters.

8.2 The effect of presentation order in pairwise choices

Keppel and Wickens wrote in their book that randomization could effectively minimize the contextual artifacts in an experiment [103] and introduced a *mixed-factor* design[1] , where double-randomization on both the stimuli and the subject could reduce the contextual effect. However, when an experiment involves multiple independent variables, it faces the situation where two solutions are possible; one may offer a completely randomized sequence of the stimuli regardless of the difference residing in each independent variable or one may conduct the randomization sequence only within a variable. Originally the dissertation research was initiated with the motivation to investigate how a context, differentiated by musical selection, would influence the listener preference when solo piano music was captured and reproduced via multichannel speakers. The study involves two main independent variables, microphone technique and musical selection, which makes it possible to hypothesize that the randomization sequence, or the presentation order, of the stimuli might affect the listeners' affective response. This issue actually has been investigated in depth and the results were published by Martens *et al.* [65], summary of which will be introduced in the following paragraphs.

In order to investigate the effect of the presentation order, two groups of subjects were formed, each containing 18 listeners, and these two groups completed trials according to two different trial ordering schemes as mentioned in the section 5.3.3 on

[1] See the Chapter 6.1 of [28] for more information about the *mixed-factor* design.

the page 68. The given task was identical to both groups so that the listeners compared two randomly selected multichannel piano music and made a choice of which they prefer. The difference between two groups is that for one group all trials for a given musical selection were completed in a single block, and then the experiment progressed to a block of trials for a different musical selection. This approach to trial ordering has been termed the *successive-treatment design*. In contrast, the second group of 18 listeners also completed four blocks of 12 preference-choice trials, but the musical selection was randomly assigned from trial to trial, so that the presentation of the four musical selections was distributed throughout the 48 trials. This approach to trial ordering has been termed the *intermixed-treatment design*.

Figure 8–1, adapted from [65] displays the estimated preference scales of the most highly preferred microphone techniques, Fukada Tree as 🄵 and Polyhymnia Pentagon as Ⓟ. Preference choices were modulated more significantly by musical selection for the group of subjects assigned to the *intermixed-treatment* condition. Whereas listeners receiving *successive* trial ordering could become acclimated to a given musical selection, and make preference choices based mostly upon the particular differences between microphone techniques, listeners receiving *intermixed* trial ordering were not given the chance to become acclimated to each musical selection, and therefore musical selection differences were more influential on preference choices for these listeners. Further, it was found in an analysis of intransitivities that the consistency of listeners who received blocks of *intermixed* trials differed from that of listeners who received *successive* trials in a manner that was consistent with differences in preference scale values, as shown in the Figure 3 of [65]. In most cases,

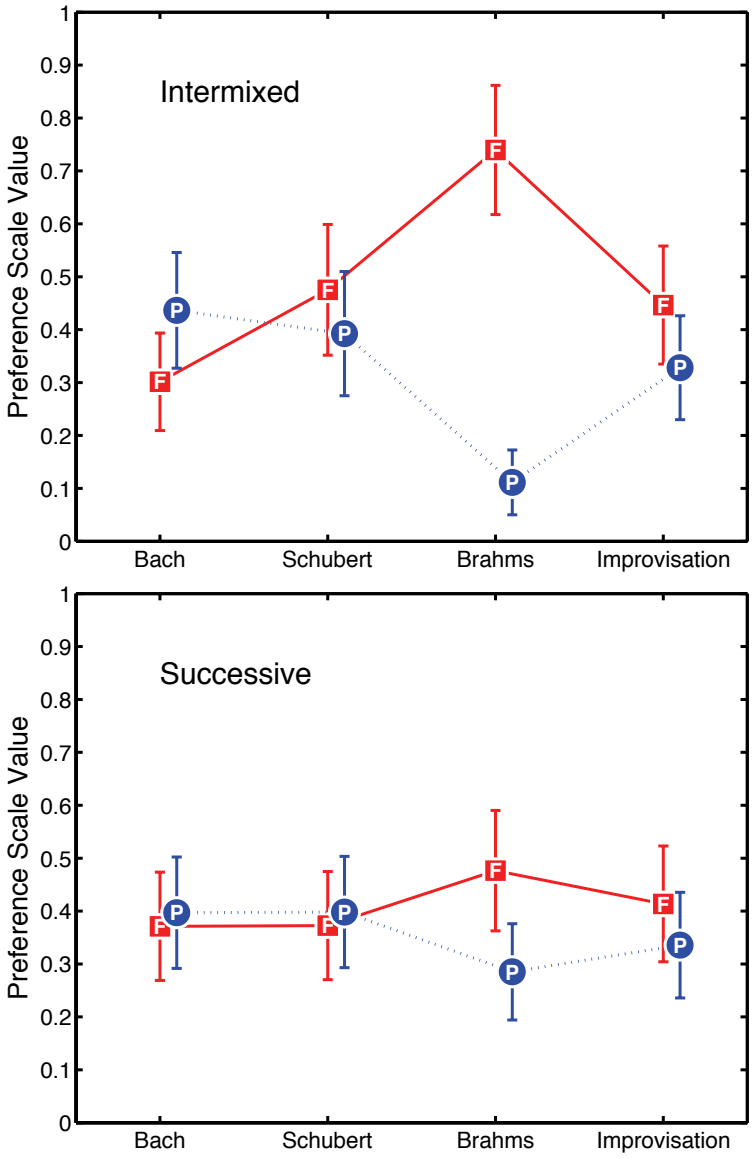


Figure 8–1: **[Upper panel]** Estimated preference scale values of the l8 listeners in the *successive-treatment* condition for imagery associated with four piano pieces recorded using two of the four microphone techniques included in the test, Fukada Tree and Polyhymnia Pentagon (plotted as Ⓕ and Ⓟ respectively). **[Lower panel]** The analogous results for the 18 listeners in the *intermixed-treatment* condition. In both of these graphs, error-bars represent corresponding 95% confidence intervals for the computed preference scale values for the two microphone techniques.

blocks of intermixed musical selections gave rise to greater inconsistency in pairwise preference choices, at least when listeners were relatively indifferent to the top two choices of microphone techniques used in the recordings. On the other hand, for the case in which stimuli differed more widely in preference, intermixing selections within a block of trials seems to have made it easier for listeners to maintain consistency. This showed that individual behavior of the listeners has been affected by the different presentation order. It might be said, therefore, that a common conclusion regarding contextual dependencies can be drawn here, regardless of whether the analyses were focused upon group behavior, or upon individual behavior.

8.3 The effect of sliding internal reference in the measure of an auditory attribute

While an affective response such as the listener's preference is more vulnerable to contextual and non-experimental variables, a measurement of a percept, if it is unidimensional and has a reference, has been known to be less influenced by the context. Martin and Bech [35] asserted that a perceptual attribute could be "objectively" measured and can be "verified externally with a different procedure." For example, a subject can report the relative magnitude of sweetness of a coffee consistently in various test conditions. However, this statement might be only assertible when a proper external reference exists; as for the coffee example, black coffee can serve as an external reference of sweetness. Many current research projects, therefore, adapt the MUSHRA (MUltiple Stimulus with Hidden Reference and Anchor) [85] method, where a *reference* and an *anchor* are presented with the test stimuli to listeners. However, it might be impossible or at the least difficult to present an external reference for the stimuli in a test, if a percept is newly identified during the experiment. In

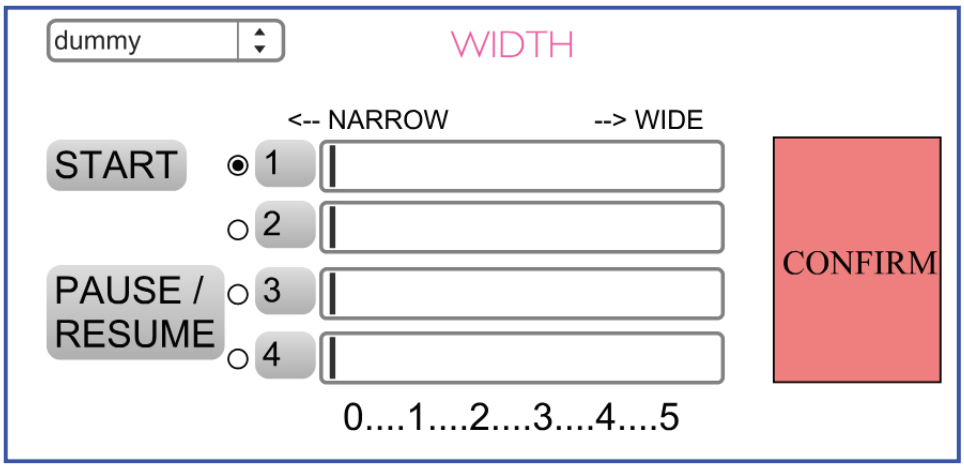


Figure 8–2: The GUI presenting multiple versions of multichannel piano sound and collecting the listeners' perceptual response of the given attribute, in this case auditory source width

particular, this dissertation research aimed to investigate the affective response of a multichannel piano sound captured by four multichannel microphone techniques. For these stimuli, it was relatively hard to create the multichannel piano sound field that can server as an external reference across various musical selections. Consequently, this dissertation research has adapted a multi-stimuli comparison without a reference to measure the relative strength of the salient percepts for multichannel reproduced piano sound.

Figure 8–2 shows the GUI that has presented the multiple stimuli differentiated by types of microphone techniques and collected the ratings of the associated attributes (source width in this case). This measuring task is confined to an identical musical selection, which required a listener to judge the perceived difference within

a musical selection. In general, if an experiment involved multiple independent variables, it would be even harder to have an external reference that is effective across the variables; the current experiment includes "musical selection" as an independent variable. In this condition, listeners would tend to apply their own internal reference derived from their memories and experiences. In other words, listeners had to deal with unfamiliarity with units of magnitude when a new musical selection was presented. For some trained listeners, it might be possible to have consistent internal reference over different contexts. One good example of such training would be the Timbral Ear Training (TET) [119] offered to master students in the Sound Recording program at McGill University. With this training, listeners can develop the ability to detect the change in the spectrum and identify the frequency and magnitude regardless of different sound sources. In contrast, there are certain auditory attributes, to which listeners find it hard to apply a constant reference over various contexts, even with enough experience. For example, a listener could estimate loudness by stating that "A is two times louder than B" or "A is not equally as loud as B" after comparing two stimuli. However, if a listener were asked to give a numeric value of loudness to A or B in PHON or SONE, one might find the task hard, thus answer inconsistently.

This is particularly problematic when the physical characters are related to the perceptual measures, since a physical quantity is often measured without accounting for the given context. In other words, mapping a set of perceptual measurements obtained from multiple conditions to a single physical metric might fail if the context

would create a non-ignorable effect on the listeners response by shifting their internal reference.

Throughout the dissertation research, predicting variance in the measurements of perceptual attributes has been one of three major tasks. As shown in Chapter 7, several prediction models were quite successfully able to account for the variance in the attribute ratings. For example, the physical measure of ASW was shown to be robust against variance between the musical programs that were presented for multi-stimulus comparisons. In particular, physical measures based upon binaural responses to a variety of multichannel reproduced piano programs were quite successful over the entire set of programs, with no apparent contextual dependence. In contrast, the previous experiment showed that a model attempting to relate physical measures of sharpness to sharpness ratings (rating here is used as an equivalent terminology of perceptual measure, in order not to confuse with the double use of the word "measure" for both the physical and the perceptual. Thus, measure without a specific description refers to a physical measure while rating refers to a perceptual measure) was not nearly as successful when the ratings were expressed as standardized scores across the entire set of stimuli. The used sharpness model, proposed by Marui and Martens [2], modified the previously well-accepted sharpness prediction model of Zwicker [29] by combining it with the product of the Zwicker Sharpness (ZS) and spectral variance of a stimulus; and successfully predicted the perceived sharpness of broadband noises in their experiments.

Hence, it was hypothesized that listeners might adjust their internal reference of auditory sharpness to the given musical selection, producing weak prediction results.

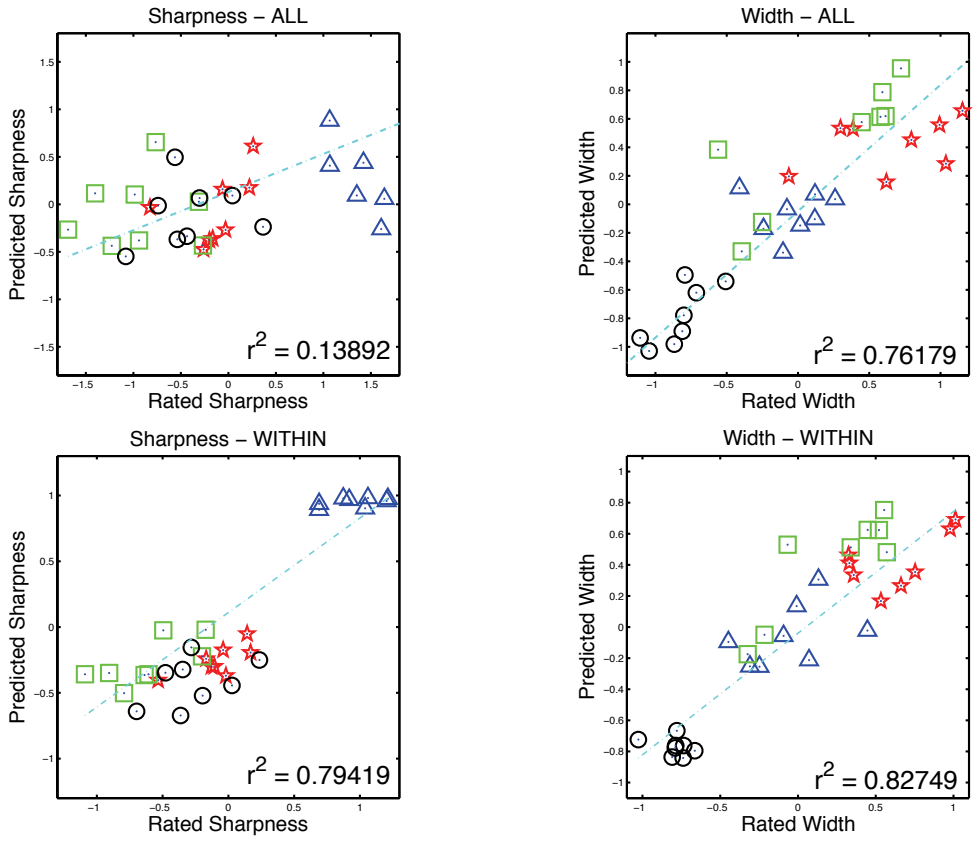


Figure 8–3: **[Left upper panel]** The result of the regression between the averages of obtained sharpness ratings centered across all stimuli (*All-Standardization*) and the physically measured sharpness, based on Marui and Martens [2] **[Left lower panel]** The result of the regression between the sharpness ratings centered across within each musical selection, i.e. four versions of same performance, (*Within-Standardization*) and the associated physical measures. **[Right panels]** The regression results of the apparent source width (ASW) ratings for *All-Standardization* and *Within-Standardization* respectively.

To test this hypothesis, a revised model of analysis was proposed; instead of relating the averaged sharpness ratings of **all** stimuli to their physical measurements, four ratings made for four microphone techniques of **each** musical selection were selected and mapped to the associated physical measures. The former method is termed *All-Standardization* because an averaged sharpness rating centered across all ratings, then these all ratings were mapped to the physical measures. Meanwhile, the latter is termed *Within-Standardization* which centered sharpness ratings "within" each multi-stimuli trial (i.e., relative to the mean sharpness for the four versions of a single musical selection being presented for comparison), then mapped to the physical measures of those four stimuli.

The two left-side panels of Figure 8–3 show the results of two analyses; the physical measures related via the *Within-Standardization* method accounted for about 80% of the variance in sharpness ratings while the former model (*All-Standardization*) accounted for only 14%. This result supports the hypothesis that listeners shifted their internal reference of auditory sharpness across different musical selections. While listeners seemed to adjust their internal reference for auditory sharpness ratings within each musical selection, they made relatively consistent ASW ratings across multi-stimulus trials as seen in the two right-side panels of Figure 8–3. When context was taken into account for ASW ratings (i.e., measured using the *Within-Standardization* method), the new approachs result accounted for only about 6% more variance than did the previous *All-Standardization* method. Auditory sharpness ratings seemed to be intrusive due to the given stimuli, while ASW was unintrusive. This is probably because listeners were able to match ASW to an external reference for width, such as

the distance between left and right speakers, whereas they failed to apply a constant reference of sharpness throughout distinctive musical selections.

It would be possible to remove such a context dependency in three ways as mentioned previously: the presence of the external reference could eliminate the observed contextual dependency; generating a set of stimuli fully randomized across all independent variables could also be a solution, whereas this might produce other arguments; and the final possible solution might be training listeners so that they can build up a more consistent internal reference of a percept. Before closing this section, a reader should be reminded that while the new Within-Standardization revealed the context dependency, the quantitative relation between physical measures and perceptual ratings was not yet experimentally verified. In order words, the method adapted herein was manipulating the "physical values" to fit the "perceptual ratings," which only supported that the rating procedure was inappropriate and that new ratings obtained with the reference or enough training of the listeners "may" be well associated with the physical measures.

CHAPTER 9
Conclusions and future work

9.1 Conclusions

Determining a recording technique that features a satisfying sound quality is a major task in classical music recording. In particular, it is not easy to find the best position of a microphone array to capture the performance of the solo piano with regard to its acoustical radiation pattern and the effect caused by the subtle interaction between the performer and the recording venue. Experienced recording engineers should have a strong internal standard of how to capture the performance of the piano considering the given acoustical and electro-acoustical conditions. The know-how of a legendary engineer or producer, unfortunately, is delivered to his/her apprentice esoterically, with few supportive theories and external validations. There have been requests to develop a method for a systemic evaluation of the multichannel piano sound in the recording process.

This research investigated the important factors characterizing the perceptual differences between the sophisticated multichannel microphone techniques in order for a preferred multichannel piano recording. Further, it aimed to experimentally devise a quantitative model that accounts for the variances in the sound quality of the multichannel piano sound. Sound quality is a broad concept, the definition of which requires clarification by the author. As stated in the first manuscript (shown in the chapter 5.3.3 on page 68), sound quality in this dissertation refers to the listener's

preference that they would rather choose and listen to a sound field captured by a multichannel microphone technique for an extended period of time in their home over loudspeakers than others. In order to serve as stimuli for the research, four multichannel microphone techniques were selected and placed in a hall to capture the identical performance of a solo piano. The purpose of the dissertation research was then achieved through the following several experiments: (1) investigating the listeners' preference; (2) predicting the obtained preference through instrumental measurement; (3) identifying and measuring the perceptual attributes; and (4) relating the quantified percepts to the associated physical measures.

This study first derived scales of the listeners' preference for the multichannel microphone techniques, and showed that the scales were modulated, not only by the microphone techniques, but also by the musical content. The obtained preference scales were also affected by the presentation order of the stimuli. Secondly, the binaurally captured signal of the multichannel reproduced piano sound were analyzed, showing that two instrumental measures - ear signal incoherence (ESI) and side bass ration (SBR) - could account for about 82% of the obtained the listeners' preference regardless of the presentation order. Thirdly, the subsequent experiment elicited the five salient perceptual attributes characterizing the used microphone techniques through a triadic comparison ([127, 128]). These attributes were then rated by selected trained listeners and the analysis of the attribute ratings (including preference rating) showed that three salient attributes - auditory source width (ASW), auditory sharpness, and bass tightness - were salient to account for the listeners' preference [86]. Finally, the variances of the magnitude of these attributes were explained by

the conventional and newly-proposed physical measurements. Further, these physical measurements accounted for the variance in the listeners' preference in a similar way that the associated attributes did. As a general conclusion, it can be said that the variances in ASW and sharpness have caused the modulation of the overall listeners' preference for the multichannel reproduced piano sound and, therefore, the preferred multichannel piano sound could be achieved through optimizing the two perceptual parameters by manipulating the associated physical quantities.

9.2 Discussion

It is worth discussing whether the overall quality of reproduced sound depends more upon timbral fidelity or more upon spatial fidelity, and in addition, which of these two accounts for more of the variance in quality for particular sets of sound sources. A recent investigation by Rumsey *et al.*. reported the relative influence of spatial versus timbral fidelity on Basic Audio Quality (BAQ) [129]. Whereas a comparative study between multichannel and other reproduction formats (such as [26]) found spatial attributes to be prominent factors that can differentiate various reproduction systems, Rumseys study showed that the overall quality is more dependent on the variance of timbre, at least for the sound sources that they included in their tests. A similar result was observed in the research presented in this dissertation; and that result can be summarized as follows: timbral fidelity can be associated with the magnitude of the first principal component of the attribute ratings, and spatial fidelity, in particular the frontal spatial image fidelity, can be associated with the second principal component of the attribute ratings. It is true

that spatial characteristics are generally enhanced in multichannel sound reproduction. However, when a listener faces two multichannel sound fields that both have satisfying spatial attributes, he/she tends to be attentive to the differences in other fundamental attributes, those being related to timbral fidelity. It does not imply that timbral fidelity will always be a superior predictor of quality ratings, nor necessarily more important than spatial fidelity; rather it reflects the fact that both fidelities are necessary in order to recreate a pleasing multichannel sound field, and a listener tends to be more attentive to what is lacking.

It should be remembered here that each stimulus used in this dissertation research project was created in order to maintain high acceptability; each recording featured an acceptable multichannel piano sound to the listener. The spatial attributes could be well reproduced via multichannel reproduction here, and so the variance in spatial fidelity perhaps modulated less the overall quality of the multichannel piano sound. Another reason for generally weak influence of spatial fidelity on variance in the overall quality evaluation across a number of studies is that spatial fidelity is quite dependent on what instruments or musical performance is being recorded and reproduced. The following gives a good example of why spatial attributes cannot be more critically important as predictors of overall quality: a comparison between a solo violin and the full orchestra will require different weights on different attributes in the prediction of sound quality, the wider ASW could positively influence on the overall quality of the multichannel reproduced orchestra music; in contrast, the same ASW may negatively affect the solo violin music. Listeners appreciate the magnitude of ASW according to the sound source. Similarly, the bass tightness might not be a

critical contributor to the quality of a string quartet, which might not able to produce enough low frequency energy to evoke the listener to sense its perceptual effect. A recent study of a comparison of microphone techniques for the orchestra by the AES Japan surround study group (part of their research result has been reported in [130]) produced different lists of salient attributes for their stimuli, which are distinctively different from the current research.

Another important topic to discuss is whether the way that test subjects listen in an experiment is similar enough to the way in which people typically listen to music for everyday enjoyment. In detail this question can be stated as following a hypothetical question: Can listeners evaluate the sound field by listening in the same manner as when they are enjoying their favorite tunes in their home or live concert? Appreciating music either in a form of performed or reproduced sound is an integrative result affected by both non-cognitive and by cognitive factors, such as memory, emotion, experience, etc. This study has assumed that these factors can be decomposed into salient perceptual attributes and contextual effects. While this has been successfully achieved in other sensory studies, especially those in food science, the listening process might be different. Interestingly, there is a point of view in behavioral psychology, for example the Gestalt theory, asserting that a whole is different than the sum of its part. Any listener appreciates the musical as a whole without considering the details of its component. Nonetheless, when listeners are asked to compare various sound fields, they naturally tend to be analytical which might cause to change the way of listening. Analytical music listening can be hard work, and as it gets harder, it might prevent people subjects listening to the music as they normally

would. To date, there has not been a report on an investigation of the relation between integrative and analytical listening. Therefore, it might be risky to solely rely on the current perceptual sound quality evaluation method, which forces to concentrate and detect the small differences. However, Olive [84] stated that the untrained listeners showed the same pattern of results as a small number of trained listeners tested similarly (of course, those subjects are expected to be listening analytically within an experimental environment). While it contained a very encouraging result, there is no guarantee that those untrained subjects listened to music as they did in their home, especially due to the fact that the experiment was done in a laboratory. If an experiment can collect data on a subjects affective response without requiring the reorientation of his/her attention, it might be possible to supply a fair ground to compare the difference between attentive (or analytical) and unconscious (or integrative) listening. Nevertheless, one thing that might legitimize the current research with respect to this issue is that the subjects were trained recording engineers who have been required to be both analytical and integrative in evaluating the sound field. For this group of listeners, even though being analytical could be relatively demanding, they nonetheless should be able to be integrative and analytical at the same time.

9.3 Future Work

The author plans to continue with new research based on the results of the research presented in this dissertation: (1) validation of the proposed prediction model to an arbitrary multichannel piano sound and corresponding calibration of the model; (2) development of a new multichannel microphone technique that can feature

the required characteristics of multichannel piano music, and of other applications; (3) investigation of the relevance between the five (or five-point-one) raw signals and their corresponding binaural signals with regard to their effect on the salient perceptual attributes, so as to develop a new quality prediction model via analysis of the five raw signals; and (4) investigating the influence of the acoustical conditions on the multichannel microphone technique. At the same time, the new research will also focus on the direction of development of an auditory imagery control system that will capture the three-dimensional information and deliver it through an ITU standard reproduction system.

CHAPTER 10
Contributions of Authors

Publication I

Paper Authors - Sungyoung Kim, Martha DeFrancisco Kent Walker, Atsushi Marui, and William L. Martens

Paper Title - An examination of the influence of musical selection on listener preferences for multichannel microphone technique

In this paper, I was the primary author.

Martha DeFrancisco contributed to achieving a pleasing sound quality on the multichannel recordings used as the stimuli in the experiment.

Kent Walker assisted the recording and the design of the listening experiment by calibrating listening conditions.

Atsushi Marui assisted the design of the graphic user interface for the listening experiment and also contributed to the statistical analysis of intrasitivities of the listeners.

William L. Martens contributed to clarification of the research questions and assisted analysis of behavioral data obtained in the experiment.

Publication II

Paper Authors - Sungyoung Kim, Atsushi Marui, and William L. Martens

Paper Title - Predicting listener preferences for surround microphone technique through binaural analysis of loudspeaker-reproduced piano performances

In this paper, I was the primary author.

Atsushi Marui assisted the analysis of the binaurally captured multichannel piano sound.

William L. Martens contributed to clarification of the research questions and assisted analysis of the obtained binaural characters.

Publication III

Paper Authors - Sungyoung Kim and William L. Martens

Paper Title - Deriving physical predictors for auditory attribute ratings made in response to multichannel music reproductions

In this paper, I was the primary author.

William L. Martens contributed to clarification of the research questions and assisted analysis of behavioral data obtained in the series of experiments reported.

Errata

- In page 108, the sentence "three subjects, S1, S3, and S5, were separated out from the other five ..." should be corrected as "three subjects, S1, ***S2***, and S5, were separated out from the other five ...".

- In page 112, two equations 7.2 and 7.3 should be corrected respectively as below:

$$w = \sum_{i=1}^{25} \mathbf{(1 - f(i))} \cdot g(i)$$

$$Bt = \sum_{i=1}^{9} \mathbf{(1 - f(i))} \cdot g(i)$$

Appendix A

Five sets of bipolar adjectives anchoring the associated salient attributes of multi-channel reproduced piano music and their definitions:

Wide ↔ **Narrow**: The apparent horizontal spatial extent of the sound source (piano). Synonyms used by listeners included: Spread / Broad ↔ Centered.

Sharp ↔ **Dull**: The variation in tone coloration associated with an increase in high-frequency content relative to low-frequency content in the sound source (piano). Synonyms used by listeners included: Shinny / Bright ↔ Dark.

Focused ↔ **Diffused**: The apparent integration of the sound source (piano) into a single unified image. Synonyms used by listeners included: Clear / Defined ↔ Blurry / Washed.

Tight Bass ↔ **Muddy Bass**: The apparent integration of the low-frequency content in the sound source (piano) into a single unified image. Synonyms used by listeners included: Solid Spectrum / Natural Bass / Focused Bass ↔ Thin Spectrum / Boomy Bass.

Distant ↔ **Close**: The apparent spatial distance of the sound source (piano) from the listening position. Synonyms used by listeners included: Far / Further ↔ Near.

Appendix B - Compliance Certificate

The certificate of the ethics review is attached at the end of this dissertation

References

[1] Søren Bech. Methods for subjective evaluation of spatial characteristics of sound. In *Proc. Audio Engineering Society 16*$^{\text{th}}$ *Int. Conf. on Spatial Sound Reproduction*, Rovaniemi, Finland, March 1999. AES.

[2] Atsushi Marui and William L. Martens. Predicting perceived sharpness of broadband noise from multiple moments of the specific loudness distribution. *J. Acoust. Soc. Amer.*, 119:EL7, 2006.

[3] English language Wikipedia. Online definition. http://en.wikipedia.org/wiki/Stereophonic_sound. December 17th, 2008.

[4] ITU-R. *Recommendation BS.775-1, Multi-Channel Stereophonic Sound System with or without Accompanying Picture.* Int. Telecommunications Union Radiocommunication Assembly, Geneva, Switzerland, 1992 – 1994.

[5] Martha DeFrancisco. Personal communication, 2007.

[6] Francis Rumsey. *Spatial Audio.* Music Technology Series. Focal Press, Oxford, UK, 2001.

[7] Michael A. Williams. Multichannel sound recording practice using microphone arrays. In *Proc. Audio Engineering Society 24*$^{\text{th}}$ *Int. Conf. on Multichannel Audio: The New Reality*, Banff, Canada, June 2003. AES.

[8] Michael A. Gerzon. General Metatheory of Auditory Localization. In *Proc. Audio Engineering Society 92*$^{\text{nd}}$ *Int. Conv.*, Vienna, Austria, March 1992. AES. Preprint 3306.

[9] Michael A. Gerzon and Geoffrey J. Barton. Ambisonic Decoders for HDTV. In *Proc. Audio Engineering Society 92*$^{\text{nd}}$ *Int. Conv.*, Vienna, Austria, March 1992. AES. Preprint 3345.

[10] Florian Camerer and Christian Sodl. *Classical Music in Radio and TV - a Multichannel Challenge*. http://www.irt.de/wittek/hauptmikrofon/Camerer.zip, November 2001. The ORF Surround Listening Test.

[11] Akira Fukada. A challenge in multichannel music recording. In *Proc. Audio Engineering Society 19th Int. Conf. on Surround Sound*, Schloss Elmau, Germany, June 2001. AES.

[12] John Klepko. 5-Channel Microphone Array with Binaural Head for Multichannel Reproduction. In *Proc. Audio Engineering Society 103rd Int. Conv.*, New York, USA, October 1997. AES. Preprint 4541.

[13] Günther Theile. Multichannel natural recording Based on Psychoacoustic Principles. In *Proc. Audio Engineering Society 108th Int. Conv.*, Paris, France, February 2000. AES. Preprint 5156.

[14] Hyun-Kook Lee and Francis Rumsey. Investigation into the Effect of Interchannel Crosstalk in Multichannel Microphone Technique. In *Proc. Audio Engineering Society 118th Int. Conv.*, Barcelona, Spain, May 2005. AES. Preprint 6374.

[15] Geoff Martin. A New Microphone Technique For Five-Channel Recording. In *Proc. Audio Engineering Society 118th Int. Conv.*, Barcelona, Spain, May 2005. AES. Preprint 6427.

[16] Kimio Hamasaki. Multichannel Recording Techniques for Reproducing Adequate Spatial Impression. In *Proc. Audio Engineering Society 24th Int. Conf. on Multichannel Audio: The New Reality*, Banff, Canada, June 2003. AES.

[17] Joerg Wuttke. Surround Recording of Music: Problems and Solutions. In *Proc. Audio Engineering Society 119th Int. Conv.*, New York, USA, October 2005. AES. Preprint 6556.

[18] Andreas Gernemann. Stereo+C: An All-Purpose Arrangement of Microphones Using Three Frontal Channels. In *Proc. Audio Engineering Society 110th Int. Conv.*, Amsterdam, The Netherlands, May 2001. AES. Preprint 5367.

[19] Yoichi Ando. *Concert Hall Acoustics*. Springer Verlag, 1985.

[20] Michael Barron and A. Harold Marshall. Spatial impression due to early lateral reflections in concert halls. *Journal of Sound and Vibration*, 77:211 – 232, 1981.

[21] Leo L. Beranek. *Concert and opera halls: How they sound*. Acoustical Society of America, 1996.

[22] Michael Barron. Subjective study of british symphony concert halls. *Acustica*, 66:1 – 14, 1988.

[23] Michael H. L. Hecker and Newman Guttman. Survey of methods for measuring speech quality. *J. Audio Eng. Soc.*, 15(4):400 – 403, 1967.

[24] Catherine Guastavino and Brian F. G. Katz. Perceptual evaluation of multi-dimensional spatial audio reproduction. *J. Acoust. Soc. Amer.*, 116(2):1105 – 1115, 2004.

[25] Jan Berg and Francis Rumsey. Identification of Quality Attributes of Spatial Audio by Repertory Grid Technique. *J. Audio Eng. Soc.*, 54(5):365 – 379, 2006.

[26] Sylvain Choisel and Florian Wickelmaier. Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference. *J. Acoust. Soc. Amer.*, 121(1):388 – 400, 2007.

[27] Nick Zacharov and Kalle Koivuniemi. Unravelling the perception of spatial sound reproduction: Analysis & external preference mapping. In *Proc. Audio Engineering Society 111*$^{\text{th}}$ *Int. Conv*, New York, USA, May 2001. AES. Preprint 5519.

[28] Søren Bech and Nick Zacharov. *Perceptual Audio Evaluation - Theory, Method and Application*. WILEY, 2006.

[29] Eberhard Zwicker and U. Tilmann Zwicker. Audio Engineering and Psychoacoustics: Matching Signal to the Final Receiver, the Human Auditory System. *J. Audio Eng. Soc.*, 39(3):115 – 126, 1991.

[30] Tomasz Letowski. Sound quality assessment: concepts and criteria. In *Proc. Audio Engineering Society 87*$^{\text{th}}$ *Int. Conv.*, 1989. Preprint 2825.

[31] Jens Blauert and Ute Jekosch. Concept Behind Sound Quality: Some Basic Considerations. In *Proc. of The 32*$^{\text{nd}}$ *Int. Congress and Exposition on Noise Control Engineering*, Seogwipo, Korea, August 2003.

[32] Francis Rumsey. Spatial Quality Evaluation for Reproduced Sound : Terminology, Meaning, and a Scene-Based Paradigm. *J. Audio Eng. Soc.*, 50(9):651 – 666, 2002.

[33] Harry T. Lawless and Hildegarde Heymann. *Sensory evaluation of food: Principles and practices*. Kluwer Academic Publishers, New York, USA, 1999.

[34] Herbert Stone and Joel L. Sidel. *Sensory evaluation practices*. Academic Press, 3rd edition, 2004.

[35] Geoff Martin and Søren Bech. Attribute Identifiation and Quantification in Automotive Audio - Part I: Introduction to the Descriptive Analysis Technique. In *Proc. Audio Engineering Society 118th Int. Conv.*, Barcelona, Spain, May 2005. AES. Preprint 6360.

[36] Sean E. Olive. A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part I - Listening Test results. In *Proc. Audio Engineering Society 116th Int. Conv.*, Berlin, Germany, May 2004. AES. Preprint 6113.

[37] Sean E. Olive. A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part II - Development of the Model. In *Proc. Audio Engineering Society 117th Int. Conv.*, SanFrancisco, USA, October 2004. AES. Preprint 6190.

[38] Sylvain Choisel and Florian Wickelmaier. Relating auditory attributes of multichannel sound to preference and to physical parameters. In *Proc. Audio Engineering Society 120th Int. Conv.*, Paris, May. 2006. Preprint 6684.

[39] Russell Mason, Natanya Ford, Francis Rumsey, and Bart De Bruin. Verbal and Nonverbal Elicitation Techniques in the Subjective Assessment of Spatial Sound Reproduction. *J. Audio Eng. Soc.*, 49(5):366 – 384, 2001.

[40] Allan Paivo. The Relationship between Verbal and Perceptual Codes. In *Handbook of Perception*, volume VIIII. Academic Press, 1978.

[41] Sylvain Choisel and Karin Zimmer. A Pointing Technique with Visual Feedback for Sound Source Localization Experiments. In *Proc. Audio Engineering Society 115th Int. Conv.*, New York, USA, October 2003. AES. Preprint 5904.

[42] Natanya Ford, Francis Rumsey, and Bart de Bruyn. Graphical elicitation techniques for subjective assessment of the spatial attributes of loudspeaker reproduction - A pilot investigation. In *Proc. Audio Engineering Society 110*$^{\text{th}}$ *Int. Conv*, Amsterdam, The Netherlands, May 2001. AES. Preprint 5388.

[43] Natanya Ford, Francis Rumsey, and Tim Nind. Communicating listeners' auditory spatial experiences: A method for developing a descriptive language. In *Proc. Audio Engineering Society 118*$^{\text{th}}$ *Int. Conv*, Barcelona, Spain, May 2005. AES. Preprint 6481.

[44] John Usher and Wieslaw Woszczyk. Design and testing of a graphical mapping tool for analyzing spatial audio scenes. In *Proc. Audio Engineering Society 24*$^{\text{th}}$ *Int. Conf. on Multichannel Audio: The New Reality*, pages 157 – 170, Banff, Canada, June 2003. AES.

[45] John Usher and Wieslaw Woszczyk. Visualizing auditory spatial imagery of multi-channel audio. In *Proc. Audio Engineering Society 116*$^{\text{th}}$ *Int. Conv.*, Berlin, Germany, May 2004. AES. Preprint 6054.

[46] Stanley S. Stevens. On the psychological law. *Psychological Review*, 64(3):153 – 181, 1957.

[47] René V. Dawis. Scale Construction. *Journal of Counseling Psychology*, 34(4):481 – 489, 1987.

[48] Eberhard Zwicker and Hugo Fastl. *Psychoacoustics - Facts and Models*. Springer, 3$^{\text{th}}$ edition, 2007.

[49] George A. Gescheider. *Psychophysics: The Fundamentals*. Lawrence Erlbaum Associated, 3$^{\text{rd}}$ edition, May 1997.

[50] Sean E. Olive, Peter L. Schuck, James G. Ryan, Sharon L. Sally, and Marc E. Bonneville. The Detection Threshold of Resonances at Low Frequencies. *J. Audio Eng. Soc.*, 45(3):116 – 128, 1997.

[51] William L. Martens, Jonas Braasch, and Wieslaw Woszczyk. Identification and discrimination of listener envelopment percepts associated with multiple low-frequency signals in multichannel sound reproduction. In *Proc. Audio Engineering Society 117*$^{\text{th}}$ *Int. Conv.*, San Francisco, CA, USA, October 2004. AES. Preprint 6229.

[52] Sungyoung Kim, William L. Martens, and Atsushi Marui. Discrimination of auditory source focus for musical instrument sounds with varying low-frequency cross correlation in multichannel loudspeaker reproduction. In *Proc. Audio Engineering Society 119*th *Int. Conv.*, New York, USA, October 2005. AES. Preprint 6544.

[53] Roger R. Davidson and Peter H. Farquhar. A bibliography on the method of paired comparisons. *Biometrics*, 32(2):241 – 252, June 1976.

[54] Louise L. Thurstone. A law of Comparative Judgment. *Psychological Review*, 101(2):266 – 270, 1994.

[55] Patric Susini, Stephen McAdams, and Suzan Winsberg. A multidimensional technique for sound quality assessment. *Acoustica*, 85:650 – 656, 1999.

[56] Stefan Brachmanski. Subjective assessment of quality of multimedia signals by means of A-B test. In *Proc. Audio Engineering Society 122*nd *Int. Conv.*, Vienna, Austria, May 2007. AES.

[57] Reiko Okumura, Kimio Hamasaki, and Kohichi Kurozumi. Distance perception of phantom sound images presented by multiple loudspeakers placed at different distance in front of listener. In *Proc. of Audio Engineering Society 121*st *Int. Conv.*, San Francisco, USA, October 2006. AES. Preprint 6891.

[58] William L. Martens and Atsushi Marui. Psychophysical calibration of sharpness for multiparameter distortion effects processing. In *Proc. of Audio Engineering Society 114*th *Int. Conv.*, Amsterdam, The Netherlands, May 2003. AES. Preprint 5739.

[59] Roy Irwan and Ronald M. Aarts. Two-to-Five channel sound processing. *J. Audio Eng. Soc.*, 50(11):914 – 926, 2002.

[60] Ralph Allan Bradley. Rank analysis of incomplete block designs: III. some large-sample results on estimation and power for a method of paired comparisons. *Biometrika*, 42(3/4):450 – 470, December 1955.

[61] Ralph Allan Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324 – 345, December 1952.

[62] Karin Zimmer, Wolfgang Ellermeier, and Christian Schmid. Using Probabilistic Choice Models to Investigate Auditory Unpleasantness. *Acta Acustica united with Acustica*, 90(6):1019 – 1028, November/December 2004.

[63] Florian Wickelmaier. *Indirect Scaling Methods Applied to the Identification and Quantification of Auditory Attributes*. PhD thesis, Aalborg University, 2005.

[64] Florian Wickelmaier and Christian Schmid. A matlab function to estimate choice model parameters from paired-comparison data. *Behavior Research Methods, Instruments, & Computers*, 26(1):29 – 40, 2004.

[65] William L. Martens, Atsushi Marui, and Sungyoung Kim. Investigating Contextual Dependency in a Pairwise Preference Choice Task. In *Proc. Audio Engineering Society 28th Int. Conf. on The Future of Audio Technology - Surround Sound and Beyond*, Piteå, Sweden, June 2006. AES.

[66] Thomas Sporer. International standard for sound quality evaluation. In *Proc. of Spatial Audio & Sensory Evaluation Techniques*, April 2006.

[67] Stanley S. Stevens. *Handbook of Experimental Psychology*, chapter Mathematics, measurement and psychophysics, pages 1 – 49. Wiley, New York, 1951.

[68] ITU-R. *Recommendation BS.1284, Methods for the Subjective Assessment of Sound Quality - General Requirements*. Int. Telecommunications Union Radiocommunication Assembly, 1998.

[69] Jum. C. Nunnally and Ira H. Bernstein. *Psychometric theory*. McGraw-Hill, 3^{rd} edition, 1994.

[70] Gösta Ekman and L. Sjöberg. Scaling. *Annual Review of Psychology*, 16:451 – 474, 1965.

[71] Ann C. Noble. Instrumental analysis of the sensory properties of food. *Food Technology*, 29(11):56 – 60, 1975.

[72] Masayuki Morimoto and Z. Maekawa. Effects of low frequency components on auditory spaciousness. *Acustica*, 66:190 – 196, 1988.

[73] John S. Bradley and Gilbert A. Soulodre. Objective measures of listener envelopment. *J. Acoust. Soc. Amer.*, 98(5):2590 – 2597, November 1995.

[74] Russell Mason, Tim Brookes, and Francis Rumsey. Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli. *J. Acoust. Soc. Amer.*, 117(3):1337 – 1350, 2005.

[75] Masayuki Morimoto and Kazuhiro Iida. A practical evaluation method of auditory source width in concert halls. *J. Acoust. Soc. Japan*, 16(2):59 – 69, 1995.

[76] Masayuki Morimoto and Kazuhiro Iida. Appropriate frequency bandwidth in measuring interaural cross-correlation as a physical measure of auditory source width. *Acoustical Science and Technology*, 26(2):179 – 184, February 2005.

[77] Kazumi Ueda, Takuji Tanaka, and Masayuki Morimoto. Estimation of auditory source width (ASW): II. ASW for two adjacent 1/3 octave band noises with different band levels. *J. Acoust. Soc. Japan*, 18:121 – 128, 1997.

[78] Kiyoaki Terada, Hirofumi Yanagawa, Mitsuharu Takagiwa, and Manabu Fukushima. Discrimination of temporal change in transient interaural cross-correlation coefficient using band-pass-filtered noise burst convolved with simulated impulse responses. *Acoustical Science and Technology*, 26(3):289 – 291, March 2005.

[79] John Spicer. *Making sense of Multivariate data analysis*. SAGE, Thousand Oaks, USA, 2005.

[80] Thomas D. Wickens. *The Geometry of Multivariate Statistics*. Lawrence Erlbaum Associates, 1994.

[81] Nils Peters, Bruno L. Giordano, Sungyoung Kim, Jonas Braasch, and Stephen McAdams. Predicting perceived off-center sound degradation in surround loudspeaker setups for various multichannel microphone techniques. In *Proc. Audio Engineering Society 125*$^{\text{th}}$ *Int. Conv.*, San Francisco, USA, October 2008. AES.

[82] Gregory J. Sandell and William L. Martens. Perceptual evaluation of principal-component-based synthesis of musical timbres. *J. Audio Eng. Soc.*, 43(12):1013 – 1028, 1995.

[83] Sungmok Hwang and Youngjin Park. HRIR customization in the median plane via principal components analysis. In *Proc. Audio Engineering Society 31*$^{\text{st}}$ *Int.*

Conf. on New directions on high resolution audio, London, UK, June 2007. AES.

[84] Sean E. Olive. Differences in performance and preference of trained versus untrained listeners in loudspeaker tests: A case study. *J. Audio Eng. Soc.*, 51:806 – 825, September 2003.

[85] ITU-R. *Recommendation BS.1531-1, Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems*. Int. Telecommunications Union Radiocommunication Assembly, 2003.

[86] William L. Martens and Sungyoung Kim. Verbal Elicitation and Scale Construction for Evaluating Perceptual Differences between Four Multichannel Microphone Techniques. In *Proc. Audio Engineering Society 122*nd *Int. Conv.*, Vienna, Austria, May 2007. AES.

[87] Jean-Jacques Rousseau as cited by Pierre Boulez. *Orientations*. Harvard University Press, 1985.

[88] Helmut Wittek and Günther Theile. *Comparison of 7 surround main microphones*. http://www.hauptmikrofon.de/orf.htm, 2001.

[89] German language Wikipedia. Online definition. http://de.wikipedia.org/wiki/Tonmeister. April 19th, 2006.

[90] John M. Eargle, editor. *Stereophonic Techniques*. AES Anthology Series. AES, 1986.

[91] Durand R. Begault, editor. *Spatial Sound Techniques, Part I: Virtual and Binaural Audio Technologies*. AES Anthology Series. AES, 2004.

[92] Francis Rumsey, editor. *Spatial Sound Techniques, Part II: Multichannel Surround Sound Technologies*. AES Anthology Series. AES, 2006.

[93] Philipp Heck and Christoph Rieseberg. *Comparison of the four different microphone arrays*. PhD thesis, Duesseldorf University, 2001.

[94] Rafael Kassier, Hyun-Kook Lee, Tim Brookes, and Francis Rumsey. An Informal Comparison Between Surround-Sound Microphone Technqiues. In *Proc. Audio Engineering Society 118*th *Int. Conv*, Barcelona, Spain, May 2005. AES. Preprint 6429.

[95] Derek Bailey. *Improvisation: its nature and practice in music.* Moorland Publishing, 1980.

[96] George Lewis. Improvised Music After 1950: Afrological and Eurological Perspectives. *Black Music Research Journal*, 16(1):91 – 122, 1996.

[97] Michael A. Williams. Microphone arrays for natural multiphony. In *Proc. Audio Engineering Society 91*st *Int. Conv.*, New York, USA, September 1991. AES. Preprint 3157.

[98] Jean-Marie Geijsen. An invited lecture. Banff Centre, 2003.

[99] Mikkel Nymand. Introduction to microphone techniques for 5.1 surround sound. In *DPA Microphones Workshop on mic Techniques for Multichannel Audio*, Banff, Canada, 2003. Audio Engineering Society 24th Int. Conf. on Multichannel Audio: The New Reality.

[100] Eberhard Sengpiel. Decca tree recording mit neumann-druckempfngern m 50. http://www.sengpielaudio.com/ DeccaTreeRecordingM50.pdf.

[101] Günther Theile. Natural 5.1 music recording based on psychoacoustic principals. In *Proc. Audio Engineering Society 19th Int. Conf. on Surround Sound*, Schloss Elmau, Germany, June 2001. AES.

[102] Michael A. Gerzon. Ambisonics in multichannel broadcasting and video. *J. Audio Eng. Soc.*, 33(11):859 – 871, November 1985.

[103] Geoffrey Keppel and Thomas D. Wickens. *Design and Analysis: A Researcher's Handbook.* Pearson Education, Upper Saddle River, New Jersey, 4th edition, 2004.

[104] Geoff Martin. The Significance of Interchannel Correlation, Phase and Amplitude Differences on Multichannel Microphone Techniques. In *Proc. Audio Engineering Society 113th Int. Conv.*, Los Angeles, CA, USA, October 2002. AES. Preprint 5671.

[105] Geoff Martin. Interchannel interference at the listening postion in a five-channel loudspeaker configuration. In *Proc. Audio Engineering Society 113th Int. Conv.*, Los Angeles, CA, USA, October 2002. AES. Preprint 5671.

[106] Wieslaw Woszczyk. Quality assessment of multichannel sound recording. In *Proc. Audio Engineering Society 12th Int. Conf. on The Perception of Reproduced Sound*, Copenhagen, Denmark, June 1993. AES.

[107] David Griesinger. The Psychoacoustics of Listening Area, Depth, and Envelopment in Surround Recordings, and their relationship to Microphone Technique. In *Proc. Audio Engineering Society 19th Int. Conf. on Surround Sound*, Schloss Elmau, Germany, June 2001. AES.

[108] Jason Corey and Geoff Martin. Description of a 5-channel microphone technique. In *DPA Microphones Workshop on mic Techniques for Multichannel Audio*, Banff, Canada, 2003. AES. Audio Engineering Society 24th Int. Conf. on Multichannel Audio: The New Reality.

[109] http://www. surrey. ac. uk/soundrec/ias/. SPATIAL AUDIO AND SENSORY EVALUATION TECHNIQUES. Guilford, UK, April 2006.

[110] Francis Rumsey. Spatial Audio and Sensory Evaluation Techniques – Context, History and Aims. In *Proc. of Spatial Audio & Sensory Evaluation Techniques*, April 2006.

[111] Slawomir Zieliński. On Some Biases Encountered in Modern Listening Tests. In *Proc. of Spatial Audio & Sensory Evaluation Techniques*, April 2006.

[112] William L. Martens. Contextual Effects in Sensory Evaluation of Spatial Audio: Integral Factor or Nuisance? In *Proc. of Spatial Audio & Sensory Evaluation Techniques*, April 2006.

[113] Sungyoung Kim, Martha DeFrancisco, Kent Walker, Atsushi Marui, and William L. Martens. Listener preferences in multichannel audio : Examining the influence of musical selection on surround microphone technique. In *Proc. Audio Engineering Society 28th Int. Conf. on The Future of Audio Technology - Surround Sound and Beyond*, Piteå, Sweden, June 2006. AES.

[114] Russell Mason and Francis Rumsey. A Comparison of Objective Measurements for Predicting Selected Subjective Spatial Attributes. In *Proc. Audio Engineering Society 112th Int. Conv*, Munich Germany, May 2002. AES. Preprint 5519.

[115] Halliday J. H. MacFie and Duncan Hedderley. Current practice in relating sensory perception to instrumental measurements. *Food Quality and Preference*, 4:41 – 49, 1993.

[116] Sean E. Olive, Peter L. Schuck, Sharon L. Sally, Marc E. Bonneville, K. L. Momtahan, and E. S. Verreault. The variability of loudspeaker sound quality among four domestic-sized rooms. In *Proc. Audio Engineering Society 99*$^{\text{th}}$ *Int. Conv.* AES, October 1995. Preprint 4092.

[117] Norman R. Draper and Harry Smith. *Applied Regression Analysis*. John Wiley & Sons, New York, 1981.

[118] Sungyoung Kim, William L. Martens, Atsushi Marui, and Kent Walker. Predicting listener preferences for surround microphone technique through binaural signal analysis of loudspeaker-reproduced piano performances. In *Proc. of Audio Engineering Society 121*$^{\text{st}}$ *Int. Conv.*, San Francisco, USA, October 2006. AES. Preprint 6919.

[119] René Quesnel. Timbral Ear Trainer: Adaptive, interactive training of listening skills for evaluation of timbre. In *Proc. Audio Engineering Society 110*$^{\text{th}}$ *Int. Conv.* AES, 1996. Preprint 4241.

[120] Søren Bech. Selection and Training of Subjects for Listening Tests on Sound-Reproducing Equipment. *J. Audio Eng. Soc.*, 40(7/8):590 – 610, 1992.

[121] William L. Martens and Sungyoung Kim. Relating listener preferences for multichannel sound programs to salient auditory attributes and binaural stimulus measurements. In *Proc. of Inter-Noise 2006*, Honolulu, USA, December 2006. INTER-NOISE.

[122] Russel Mason, Tim Brookes, and Francis Rumsey. Integration of measurements of interaural cross-correlation coefficient and interaural time difference within a single model of perceived source width. In *Proc. of Audio Engineering Society 117*$^{\text{th}}$ *Int. Conv.*, San Francisco, USA, October 2004. AES. Preprint 6317.

[123] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and Models*. Springer-Verlag, Heidelberg, Germany, 2$^{\text{nd}}$ edition, 1999. ISBN: 3540650636.

[124] Malcolm Slaney. Auditory Toolbox: A MATLAB toolbox for auditory modeling work. Technical Report 45, Apple Computer, Apple Corporate Library, Cupertino, CA 95014, 1994.

[125] Masayuki Morimoto, Kazumi Ueda, and Masakazu Kiyama. Effects of frequency characteristics of the degree of interaural cross-correlation and sound pressure level on the auditory source width. *Acustica*, 81:20 – 25, 1995.

[126] Eustace C. Poulton. *Bias in quantifying judgments*. Lawrence Erlbaum Associates, Hove and London, UK, 1989. Christopher.

[127] Florian Wickelmaier and Wolfgang Ellermeier. Deriving auditory features from triadic comparisons. *Perception & Psychophysics*, 69(2):287 – 297, 2007.

[128] Atsushi Marui and William L. Martens. Using paired versus triadic comparison tasks as the basis for multidimensional perceptual scaling of nonlinear distortion effects for guitar timbre. In *Proc. of the 12$^{\text{th}}$ Regional Convention*, Tokyo, Japan, July 2005. AES.

[129] Francis Rumsey, Slawmir Zieliński, and Rafael Kassier. On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *J. Acoust. Soc. Amer.*, 118(2):968 – 976, 2005.

[130] Toru Kamekawa, Atsushi Marui, and Hideo Irimajiri. Correspondence relationship between physical factors and psychological impressions of mirophone arrays for orchestra recording. In *Proc. Audio Engineering Society 123$^{\text{rd}}$ Int. Conv.*, New York, USA, October 2007. AES.

McGill University

ETHICS REVIEW
RENEWAL REQUEST/FINAL REPORT

Continuing review of human subject research requires, at a minimum, the submission of an annual status report to the REB. This form must be completed to request renewal of ethics approval. If a renewal is not received before the expiry date, the project is considered no longer approved and no further research activity may be conducted. When a project has been completed, this form can also be used as a Final Report, which is required to properly close a file. To avoid expired approvals and, in the case of funded projects, the freezing of funds, this form should be returned 3-4 weeks before the current approval expires.

REB File #: 90-0907
Project Title: Investigating listeners' preferences for multichannel microphone techniques
Principal Investigator: Sungyoung Kim
Department/Phone/Email: Music Research / 398-3545 / sungyoung.kim@mail.mcgill.ca
Faculty Supervisor (for student PI): William L. Martens

1. Were there any significant changes made to this research project that have any ethical implications? ___Yes (No)
   If yes, describe these changes and append any relevant documents that have been revised.

2. Are there any ethical concerns that arose during the course of this research? ___ Yes (No) If yes, please describe.

3. Have any subjects experienced any adverse events in connection with this research project? ___ Yes (No)
   If yes, please describe.

4. (O) This is a request for renewal of ethics approval.

5. ____ This project is no longer active and ethics approval is no longer required.

6. List all current funding sources for this project and the corresponding project titles **if not exactly the same** as the project title above. Indicate the Principal Investigator of the award if not yourself.

Principal Investigator Signature: Kim Date: 2008.09.17

Faculty Supervisor Signature: William L. Martens Date: 2008.09.17
(for student PI)

For Administrative Use **REB:** ___REB-I ✓REB-II ____REB-III

____ The closing report of this terminated project has been reviewed and accepted

✓ The continuing review for this project has been reviewed and approved

✓ Expedited Review ___ Full Review

Signature of REB Chair or designate: I Mitchell Date: Apl 29, 2008

Approval Period: Oct. 2008 to Sept. 30, 2009

Submit to Lynda McNeil, Research Ethics Officer, McGill University, 1555 Peel Street, 11th floor, Montreal, QC H3A 3L8
tel:514-398-6831 fax: 514-398-4644 email:lynda.mcneil@mcgill.ca (version 02/08)

McGill

Research Ethics Board Office
McGill University
845 Sherbrooke Street West
James Administration Bldg., rm 419
Montreal, QC H3A 2T5

Tel: (514) 398-6831
Fax: (514) 398-4644
Ethics website: www.mcgill.ca/researchoffice/compliance/human/

Research Ethics Board III
Certificate of Ethical Acceptability of Research Involving Humans

REB File #: 90-0907

Project Title: Investigating listeners' preferences for multi-channel microphone techniques

Principal Investigator: Sungyoung Kim **Department:** Faculty of Music

Status: Ph.D. student **Supervisor:** Prof. W. Martens

Funding Agency and Title (if applicable): FQRSC- L'Audio haute re resolution: un outil garantit la qualité des archives musicaux pour le future

This project was reviewed on Sept. 28, 2007 by

Expedited Review ✓
Full Review ____

Mark Baldwin, Ph.D.
Chair, REB II

Approval Period: Oct. 2, 2007 **to** Oct. 1, 2008

This project was reviewed and approved in accordance with the requirements of the McGill University Policy on the Ethical Conduct of Research Involving Human Subjects and with the Tri-Council Policy Statement: Ethical Conduct For Research Involving Humans.

---

* All research involving human subjects requires review on an annual basis. A Request for Renewal form should be submitted at least one month before the above expiry date.
* When a project has been completed or terminated a Final Report form must be submitted.
* Should any modification or other unanticipated development occur before the next required review, the REB must be informed and any modification can't be initiated until approval is received.