Using Telepresence to Evaluate the Relevance of Acoustic Parameters for Musician Communication in Indeterminate Music

by

Anne Elizabeth Guthrie

A Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the degree of

MASTER OF SCIENCE IN ARCHITECTURAL SCIENCES

Major Subject: ARCHITECTURAL ACOUSTICS

Approved:

____

Jonas Braasch, Thesis Adviser

____

Paul Calamia, Thesis Co-Adviser

Rensselaer Polytechnic Institute
Troy, New York

August, 2008
(For Graduation December 2008)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENT

Thanks to Jonas Braasch for advising me in this research. Thanks also to Paul Calamia and Ning Xiang, for assisting me in my work at Rensselaer Polytechnic Institute. Thanks to Daniel Valente, Pauline Oliveros, Curtis Bahn and Bobby Gibbs for their help with setting up my experiments. Thanks to all the subjects who participated in these experiments, to Billy Gomberg for Max/MSP assistance and to Kathleen Stetson for helping me plan my analysis. Thanks also to Petr Kotik and the S.E.M. Ensemble, for continuing to champion new music and providing me with the impetus for this research.

ABSTRACT

In situations of indeterminate musical performance (particularly in telepresence, where acoustic degradation is a frequent concern), autonomous musical communication, both practical and artistic, forms the crux of the musical material. The relevance of stage acoustic and psychoacoustic parameters to contemporary performance situations have been re-examined in this research, with regard to the heightened importance of communication. Parameters developed by Gade and others are starting points for this examination. Experiments were conducted with two instrumentalists playing excerpts from a composition by Christian Wolff (open notation allows for measurable variations depending on communication quality), communicating telematically between two virtual environments. Parameters determined by questionnaire to have the strongest effect on quality and efficiency of communication were evaluated by the performers. Five parameters were tested: Self-to-Others Ratio (SOR), Reverberant Energy, Early Reflected Energy, Sound Source Position and Visual Connection. SOR was found to have the most significance for complex communication, with values as low as –12 dB preferred. The lack of Visual Connection also was considered detrimental to communication. Reverberant Energy and Sound Source Position were subject to individual variability, with trends roughly dictated by instrument. A musical analysis of performances from the Early Reflected Energy and Reverberant Energy tests showed some visible trends in the performed content, indicating that these parameters may relate to aesthetics rather than functionality.

1. Introduction

In his work "Cartesian Meditations: An Introduction to Phenomenology," Husserl describes the intentional perceptive process as a synthesis of all modes of psychic being [28, pp. 48]:

> In the case of any spatial thing, [the phenomenologist] explores its (potential and perhaps actual) changing perspectives; furthermore, with regard to its temporal modes of givenness, the modifications of its being still intended while it sinks retentionally into the past and, with respect to the Ego, the modes of his specifically own still-having and holding, the modes of attention, and so forth [...] making present in phantasy the potential perceptions that would make the invisible visible.

In following this axiom, it becomes necessary to examine acoustics as it pertains to all classes of perception, including passive and performative, simple and complex, practiced recollection and spontaneous improvisation, objective (room acoustics) and subjective (psychoacoustics). Neuhoff, in "Ecological Psychoacoustics," describes the situation as follows [44, pp. 3]:

> One of the goals of [Ecological Psychoacoustics] [...] is to examine some of the complex interactions between divergent areas of hearing science, tying together the occurrence of acoustic events, physiological responses of the auditory system, the perceptual and cognitive experience of the listener, and how this experience influences behavior, action, and subsequent perception [...] Illuminating some of the converging evidence that is emerging from these traditionally divergent paths may yield a broader and perhaps more informative perspective on auditory function.

In the past several years, acoustics as pertains to the audience had seen rapid and extensive development, but the realm of stage acoustics has been the focus of a small percentage of this development. The seminal research into this field, conducted by Gade in the early 1980s, has provided an important resource and two significant objective acoustic parameters (Stage Support and Early Ensemble Level) for any acoustician hoping to design a stage enclosure ideal for musicians performing classical music [19]. Gade used, in particular, the Trio Sonata by J.S. Bach and the 40th Symphony by W.A. Mozart as his source material. More strictly notated and individually independent music could not have been written. In this music, ensemble playing is indeed crucial, but the necessary coordinations between musicians include stylistic and intonational coordinations, not structural or durational ones. The conductor leads the ensemble, and it is possible and at times necessary to tune out the other musicians in order to follow the score and the tempo of the conductor. In particular, the material created by the other musicians is predictable (it is already notated in the score) and an immediate, running mental analysis of the phrasing and development of all the other instrumental parts is not crucial to the creation of one's own musical material. After all, this material has already been composed. Gade found that important parameters in these situations include "hearing each other," "support," "reverberant energy," "time delay," and "timbre." Without ideal values for each of these parameters, musicians of classical music feel that their performance is compromised.

In the last 100 years, contemporary music has developed from a simple shift from tonal harmony to 12-tone serialism, into a wide field comprising both heavily notated, structured music and open improvisation. Improvisation itself has evolved from traditional jazz improvisation, where traditional jazz instruments outline a basic chord pattern and comply to a predetermined rhythm, to free improvisation, where any stylistic conceit or anarchic situation may occur, with any combination of instruments and any amount of performing experience.

Consequently, the most interesting development of contemporary music is that of indeterminate music, which can be defined as a form of open notation in which significant elements are left open by the composer, intended to be improvised or delineated by chance or other means (such as a system of rules, like a game) prior to (e.g. Cage) or during (e.g. Christian Wolff) the performance. Often these elements include stylistic, pitch, or rhythmic material, and sometimes require the musician to respond based on the surrounding sonic environment. This kind of spontaneous response requires a whole different kind of acoustic clarity than that required by classical music, but in contemporary performance practice many elements – choice of venue, rehearsal practice, seating arrangement, etc. – are highly reminiscent of classical music performance. Having worked with a contemporary orchestra for 3 years, it has become clear that these situations need to be changed.

Ideally, the findings of this research should provide contemporary musicians with a map of room-acoustic conditions favorable to indeterminate music, and with the focus on complex communication in general, a set of conditions that may be extrapolated to other forms of complex communication, such as conversational speech, political speech, audio post-production, contemporary theater, etc. The ultimate goal of this research is to distill that essence of complex communication that requires a change in the acoustic quality of a space, and to determine the ability of the current used spaces to generate this quality.

In the collection of essays, "Communication Acoustics," Blauert defines communication as "the transmission of signals carrying information across space and time" [4]. Sebastian Möller uses the term "human-to-human interaction" to describe the type of complex communication over a network that is dealt with in this research [43].

In order to determine what this essence of communication is that provides an ideal example in indeterminate music, the contemporaneous work of Jekosch, Van Valkenburg and Kubovy highlight the new forays into fields such as semio-acoustics as an attempt to describe how we process audio information and react to it in terms of quality and meaning. Semiotics may have originated in the world of psychology and speech, but it has since been expanded to understand art theory, and now psychoacoustics. Jekosch pinpoints *Gestalt* theory as the major auditory process for the signification and interpretation of sounds, and the subsequent listener reactions (in this case, the musical response) [30]. Valkenburg and Kubovy have an opposing theory, which is that *Gibsonian*, multi-modal processes are at work in these situations [66]. Both of these representations will be kept in mind when analyzing the results of the current research, in an attempt to read the acoustic requirements for indeterminate music through a broader lens of object-selection and semiotics in general. What is it about this spontaneous human-human interaction that is affected by the acoustics of a space, and what effect does this have on our interpretation of meaning and subsequent reaction?

In this research, a literature review was undertaken to understand the acoustic parameters crucial to classical music. Reviews of Gade's, and more recently Ueno's, work helped narrow the required parameters and ideal values for both solo and ensemble playing of Mozart and Bach [20,62]. Ternström's work provided values for the Self-to-Others Ratio that were common among singers in operatic performance [59]. Work by Meyer, Kirkwood, and Otondo examined the individual instrument parameters (directivity, transient spectrum) that affect room acoustics when performing [41,36,48]. Based on this review, a list of 21 parameters was developed and 9 contemporary musicians were asked to rate these parameters in order of relevance to contemporary music performance. The survey results will be listed in more detail in this work. On completion of this survey, a comparison of the ratings with the important parameters listed in classical stage acoustics works led to the selection of five parameters: Self-to-Others Ratio, Sound Source Position, Source Visibility, Reverberant Energy, and Early Reflected Energy. The values to test for these parameters were chosen based on traditional values for classical and contemporary performing spaces, influenced by information from Beranek's *Concert Halls and Opera Houses* [2], as well as the authors listed above, the constraints of the software used, and an additional factor which has not yet been addressed but is crucial to the progress of contemporary music: that real-time virtual auditory and visual venue, "telepresence."

Telepresence is rapidly becoming a popular venue for many forms of complex communication, from business and political interactions to video games to music education. Telepresence is the immersion of multiple subjects from remote locations in a single virtual environment [6]. Large acoustic firms like ARUP are already outfitting new concert halls with telepresence options (such as the concert hall for the New World Symphony in Miami), but it is a venue that is convenient and useful even in a modest lab or studio. With the proliferation of telepresence technology, much work has been done to examine the quality of the system itself, such as perceptual limens of network latency and sonic distortion, ambient noise, dropped packets and plausibility of visual/auditory asynchrony [5]. However, the options for processing the anechoic audio signal that has been transmitted from one remote location into the Virtual Auditory Environment are manifold and include real-time convolution with real impulse responses as well as the addition of synthetic reverberation or other perceptual effects.

This research makes use of the flexibility and absolute control over all aspects of the audio-visual environment provided by telepresence and utilizes it as a test environment. The audio of each musician is transmitted over the Internet and processed using spatialization software (the Virtual Microphone Control algorithm [7], explained in detail in Section 3.3.4, was readily available and determined to be ideal for this research), where the individual parameters listed above can be varied without affecting any other parameters. The goal of this research is not to find the ideal parameters for telepresence specifically, but of course any parameter values that would be ideal for a physical venue can be extrapolated easily to telepresence, and this tool allows the researcher to push the boundaries beyond physically realistic situations and find the absolute perceptual limits of certain parameters.

Additionally, psychoacoustic work by Brungart, Shinn-Cunningham and others provided insight into effects of masking between musicians and possible acoustic situations that might provide release from such masking. A detailed examination of this aspect was not undertaken in this research because of time constraints and subject availability, but the concepts are briefly addressed in the analysis.

Psychoacoustics work in auditory-stream segregation and selective attention can also be used as a filter to examine the results. When interacting with other musicians and responding creatively to unforeseen material, how to we group that material into elements we choose to incorporate into our own response? How does an acoustic degradation or enhancement shift this grouping process? These questions have been held in mind while structuring and executing this research, as shown in this paper.

Additionally, cross-modal[1] psychoacoustic research [37] has an important correlation to this research as well. There is evidence that the brain interprets information as mode-neutral (activity has been shown in the visual processing centers of the brain during the perception of auditory stimuli) or that at least there is significant bleeding between these genres of perception [54]. For this reason, the Source Visibility parameter has been tested in this research in tandem with the Sound Source Position. More detailed research into the Audio-Visual cross-modal interactions in virtual environments can be found in the Neuhoff book as well as concurrent research by Valente [65].

---

[1] With a different mode assigned to each sensory organ, this includes auditory, visual, haptic, olfactory, and gustatory modes based on the designation of their processing centers in the brain and the interaction between information processed by these different centers.

In summary, this work will outline the aforementioned literature with its relation to the topic of complex communication, then describe the subsequent experiment setup along with a brief analysis of the chosen musical material. Analysis of the subjective test results (numerical preference ratings) will then be presented, followed by an analysis of the performed content and a discussion of real-world applications. Finally, this research will attempt to answer the following question: how can we determine the acoustic quality of a space with regard to providing the best essential complex communication between human beings that fosters unique and enjoyable collaborations in a variety of venues, even over the internet?

2. Historical Review

2.1 Stage Acoustics

As mentioned previously, the work outlined in this paper has been significantly influenced by previous research in the field of Stage Acoustics. In the past 50 years, stage acoustics have taken a prominent position in acoustic design. Initially, Beranek suggested in 1962 that reverberation was important to a musician's impression of his or her own sound in a hall [3]. Subsequently, Marshall and Krokstad, among others, began to research the importance of early reflections to these soloistic impressions [39,40]. As for ensemble impressions, early reflections were credited as early as 1968, when Jordan suggested an objective stage acoustics parameter called the "Inversion Index" [32]. This parameter is simply the ratio of Early Decay Time (EDT) in the audience to the EDT onstage, based on the idea that if this ratio is larger than 1, then the early energy will build up faster onstage than in the audience, and therefore have a strong effect on support. Thomasson in 1974 also suggested another parameter, EDT5, which has been defined below [61].

Marshall tested these hypotheses in 1978 and determined that early reflections are important from 17-35 ms after the direct sound (path length of reflected sound 5.8-12 m longer than direct sound path length), and also demonstrated some frequency dependence of these reflections (high-frequency reflections are more important than those below 500 Hz) [40]. This corresponds to Meyer's hypothesis that higher frequencies (the frequencies created by transients, or attacks) are more important for rhythmic precision [41]. However, as confirmed by Meyer in 1978, due to the masking of overtones in most instruments, the masking threshold of other instruments in the presence of self-produced sound is lower at low frequencies (the specific frequency range is of course related to the range of the instrument played, and additionally, the emphasis on harmonics becomes stronger in relation to the fundamental with an increase in self-produced volume). Of course, due to instrument directivity (which narrows significantly at higher frequencies), the masking threshold of other instruments is highly dependent on the angle of incidence. For example, Meyer shows that for a violinist, at 1 kHz, the masking threshold is 10 dB lower at lateral and elevated positions than elsewhere, and at 2 kHz the threshold is lowest for frontal and elevated positions [42].

In the 1980s, Gade conducted extensive subjective experiments and determined two stage acoustics parameters that are still widely used [20]. The Stage Support (divided into three sub-parameters, early: ST1, middle: ST2, and late: ST3)[2] parameter was created to address soloist concerns, or the musician's impression of their own sound in the hall and how this relates to the aesthetics of their performance. The parameter is shown below:

$$\mathrm{ST1} = 10\log_{10}\left[\frac{\int_{20}^{100} p^2(t)dt}{\int_{0}^{20} p^2(t)dt}\right] \mathrm{dB}$$

where $p(t)$ is the sound pressure of the impulse response [2].

The impulse response is measured with a microphone 1 meter away from the source (based on the assumption that the direct-sound path-length from instrument to ear is approximately 1 m). The 100-ms cutoff time for the ST1 parameter was chosen because it corresponds to the duration of a short tone in music and also to the integration time of the ear.

The second parameter, which Gade proposed to address ensemble concerns, or musician's impressions of the relationship between their sound and the sound from the rest of the ensemble (also referred to as "functional" concerns), is called Early Ensemble Level (EEL). Although Dammerud and even Gade himself have later questioned the validity and reliability of this parameter, and ST1 has actually been shown to correlate more directly with subjective "ensemble" impressions [15,23], the experiments leading to the choice of this parameter and its definition will be described in detail below due to their influence on the current experimental design.

In 1981, Gade conducted subjective evaluations with professional classical musicians of all instrumentations, including pianists, singers, orchestral performers, and

[2] In 1989, Gade revised his ST parameters, changing the cutoff times to create the parameters $\mathrm{ST}_{\mathrm{early}}$ (20-100 ms), $\mathrm{ST}_{\mathrm{late}}$(100-1000 ms), and $\mathrm{ST}_{\mathrm{total}}$ (20-1000 ms) [22].

conductors. Based on the survey and the instrument-dependence of the results, Gade ranked the chosen parameters in order of importance and designated them based on importance for "ensemble" (functional) or "soloist" (aesthetic) playing [19]. In 1982, Gade built on this survey by testing subjective preference for specific objective parameters and using multidimensional scaling to determine the objective parameters most directly related to the abovementioned subjective list. His initial matching of subjective and objective parameters is listed in Table 1:

Table 1: Subjective-Objective Parameter Correlations in A.C. Gade [20]

1.	Reverberation (soloist concern): T_{20}, TA[3], C_{80}
2.	Support (soloist concern): ST1, ST3
3.	Timbre (soloist concern): Frequency-dependence of T_{20}, TA, C_{80}, ST1, ST3
4.	Dynamics (soloist concern): Assumed to be related to "support."[4]
5.	Hearing Each Other (ensemble concern): EDT, EDT5[5], C_{80}[6], ST1, ST3, Frequency dependence of all parameters listed
6.	Time Delay (ensemble concern): Delay of direct sound (calculated based on source-receiver distance assuming that sound travels 1 m every 2.8 ms).

---

[3] T_{20} is the time in seconds it takes an impulse to decay in a space from –5 dB to –25 dB, extrapolated out to –60 dB or multiplied by 3 [29]. TA is the decay time in seconds to –20 dB multiplied by 3.

[4] Because of the difficulty in testing such a parameter, this was not tested in the subsequent experiments.

[5] EDT (Early Decay Time) is the time in seconds it takes an impulse to decay by 10 dB, extrapolated to 60 dB, EDT5 is the time it takes to decay by 5 dB, extrapolated to 60 dB.

[6] Clarity: $C_{80} = 10\log_{10}\left[\frac{\int_0^{80} p^2(t)dt}{\int_{80}^{\infty} p^2(t)dt}\right]$ dB [2].

These parameters were measured by Gade using the impulse-response method [29], with a source and receiver onstage in different orchestral positions in 3 concert halls: Danish Radio Studio 1, Tivoli Concert Hall, and St. Anne Gymnasium Concert Hall. The soloist positions were based on the measurement techniques defined by the Stage Support parameter, and the ensemble positions were based on the distances between orchestral instrument groups (6-8m). Based on these measurements, the halls offered the ranges of parameter values shown in Table 2.

Table 2: Range of parameter values for the three measured halls, A.C. Gade

	Soloist Condition	Ensemble Condition
T_{20}	1.2-1.8 s	1.7-2.2 s
TA	0.7-1.2 s	n/a
EDT	n/a	1.2-1.8 s
EDT5	n/a	0.9-1.8 s
C_{80}	10.4-12:3 dB	–1.6-4.6 dB
ST1	–12.5- –5.5 dB	2.5-9.1 dB
ST3	–9.9 - –4.6 dB	6.6-10.5 dB

These values were then used to determine the test conditions for Gade's experiments. Experiments were also conducted to determine thresholds of audibility for a single reflection and a cluster of six reflections. The mean threshold of single reflections was –9 dB for strings and –15 dB for flutes, indicating the high level of variance between instrument groups (possibly due to masking based on instrument directivity patterns and the geography of the instrument's acoustic center in relation to the ears). A comparison of the ST1 parameter measured in the abovementioned halls shows a correlation between this audibility threshold and subjective impressions of the halls: halls with ST1 values below this threshold resulted in impressions of "lacking support."

An impulse response was simulated using a delay unit, reverberation chamber and mixer to allow more control over the individual parameters. All values for direct sound, discrete early reflections, and statistical reverberant tail were based on the expected attenuation due to spherical spreading in a hall. Then these values were changed in relation to this control value (see Table 3):

Table 3: Small Ensemble (3 musicians in a single space, rating ensemble and soloist concerns, performing *Trio Sonata #2 Trio Sonata #2* by Johann Sebastian Bach)

Early Reflected energy delay	15 ms, 32 ms, 45 ms, 60 ms, 73 ms
Early Reflected energy level	0 dB, -6 dB (normalized to control value)
Reverberant energy level	0 dB, -4 dB (normalized to control value)
High frequency early reflected energy	0 dB, -8 dB (normalized to control value)
Low frequency early reflected energy	0 dB, -8 dB (normalized to control value)

Gade found that for small ensembles, an initial reflection delay of 45 ms was preferred, but that this finding was strongly dependent on the chosen L_{refl} for the experiment. He found that raised early-reflected energy levels had a significant positive effect on feelings of support, and slight negative effect on feelings of reverberation, although the higher reverberation levels were preferred. Early reflections had a stronger effect on the preference than reverberation. He found little correlation between any parameters and impressions of "Hearing Each Other," but he found a correlation between T_{20}, TA (see Footnote 3 for definition) and "Reverberation" in the soloist concern (but less for the ensemble concern), and a correlation between ST1, ST3 and "Support" in both cases. As for instrument-dependence, violins preferred high reverberant energy levels over other instruments, and cellists preferred strong early reflections over other instruments. As for the timbral concerns, preferences were more divided: the violinists preferred the prevalence of high frequencies and flutes and cellos preferred the presence of low frequencies. The subjective impressions of "Timbre" were strongly correlated to frequency-dependent values of ST1 and ST3. In general, a high correlation was found between soloist and ensemble concerns for small ensembles.

Table 4: Large Ensembles (2 musicians separated into 2 physical spaces, performing *Symphony #40, Mvt. 3* by Wolfgang Amadeus Mozart)

Delay of direct sound of remote musician	7ms (2m), 20ms (7m), 35ms (12m), 50ms (17m), 65ms (22m), and 80ms (27m)
Ceiling reflection Delay	26ms (8.9m), 42ms (14.4m), 62ms (21.5m)
Ceiling reflection: L_{refl}	0 dB, –4 dB, –8 dB
Ceiling reflection Spectrum	Flat, High-pass filtered
Direct sound energy	0 dB (unobstructed sight), − 7 dB (obstructed sight)
Early reflected energy	±3 dB (in relation to natural spherical attenuation)
Reverberant energy	0 dB, –4 dB

Although Gade found some errant results caused by suspicion of the unrealistic environment, the results for the large ensemble tests (see Table 4) showed a negative effect of direct-sound delays longer than 35 ms (>12m). The shortest ceiling reflection delay (corresponding to highest L_{refl} was preferred, and at this high level a flat spectrum was preferred. A strong preference was shown for sound fields with high direct sound and early reflected energy level as well as lower reverberant energy levels. A strong correlation was found between "Hearing Each Other" and the ratio of sound received to sound emitted, and also high-frequency values for C_{80}, but not most parameters that compare different parts of the received sound. Because of this, the EEL parameter was proposed that compared sound at two positions, one close to the source and one at a reasonable distance for a receiver elsewhere in the orchestra. The "emitted sound" receiver is 1 m from the source and the "received sound" receiver is 8.5 m from the source. Because higher levels of reverberation were correlated slightly with lower impressions of "hearing each other," this parameter was recommended in conjunction with C_{80} and EDT measurements.

In his conclusion, Gade summarizes the relationships between subjective and objective parameters shown in Table 5.

Table 5: Summary of A.C. Gade findings

1.	Reverberation (soloist)	T_{20}, TA, C_{80}	Higher values preferred
2.	Support (soloist)	ST1	Higher values (above –10 dB) preferred
3.	Timbre (soloist)	Early Reflection Spectra	High frequencies preferred by violins, Low frequencies preferred by cellists and flutists
4.	Hearing Each Other (ensemble)	EEL, EDT, C_{80}	High values of high-frequency early energy preferred, low values of reverberation preferred
5.	Time Delay (ensemble)	Direct sound delay	Delays of less than 20 ms preferred

Ueno, et al. conducted similar experiments using a version of the Gade large ensemble setup in 2005 using real-time convolution with impulse responses from seven concert halls [63]. From these experiments, it was found that while string players preferred an early reflected energy level that is neither too weak nor too strong, wind players occasionally preferred stronger early reflections (absolute levels, not equalized to 0 dB for attenuation due to spherical spreading, ranged from –23.5 dB to –16.5 dB below the direct sound). They also found that, although the effects of reverberation time were small, an increase was generally preferred (times tested varied from 1.36 s to 2.22 s). With reverberant energy levels, a condition that was neither too weak nor too strong was preferred (conditions ranging from –23.4 to –14.8 dB below the direct sound) by strings and a stronger reverberation was preferred by winds.

Recently, Ueno et al. have examined changes of musical content in classical music performed in different halls, which is especially interesting in light of the wide variation in content from a single contemporary score [64]. Although the variations possible in classical music (one of the pieces they used was *Ave Maria*) are not as large, these experiments did find shifts in tempo, SPL, and the range of frequency and intensity vibrato for several different instruments depending on the hall. Although these experiments used autocorrelation techniques to determine the differences, something that would not be very useful in comparing contemporary music performances, the work shows that researchers are paying more attention to the content of the performance itself as a measure of the acoustics of a space, something this research also attempts to address.

The works described above are crucial to this research because they have defined what subjective and objective parameters are important to classical music and both symphonic and chamber ensemble instrumentation. The solo and ensemble comparisons are especially important because in contemporary music there is often a tenuous balance between soloist and ensemble playing that is shifted depending on the clarity of communication between musicians. A logical progression to contemporary music would involve the testing of the aforementioned parameters with different musical material and instrumentations. However, as will be described below, the introduction of new performance venues and the emphasis on different psychoacoustic principles require additional parameters to be tested beyond the six listed here.

As mentioned above, directivity of individual instruments influences the stage acoustics in complex ways. Meyer has shown how multiple directivity patterns can interfere with each other in different orchestral seating arrangements and within individual instrument groups [41]. In groups with particularly varied directivities, such as the woodwinds (flutes, bassoons, oboes, clarinets), the balance of this section changes drastically depending on the position of an audience member or a musician from another section of the orchestra. Kirkwood [36] has conducted listening experiments to determine the audibility and importance of directivity patterns (vs. an omnidirectional average) and determined (in spite of large experimental errors which prevented many precise conclusions) that loudness, reverberance, and clarity judgments were all effected by the changes in directivity patterns and that the specific directivities were preferred by the listeners to simulated omnidirectional averages.

In contemporary music, where extended techniques (a performance style that uses the instrument in an unconventional manner, creating a wide variety of timbral changes and noises, such as half-valve or multiphonic techniques) are often used, the directivity patterns may also change from one technique to another. Some preliminary measurements of French horn directivities for extended techniques completed by the author confirm this hypothesis.

Additionally, Gade addressed the issue of comb-filtering of direct and reflected sound propagating in the orchestra by using a spectral correction curve in his experiments to simulate the direct sound attenuation. This curve was determined by Krokstad [20] in experiments with rows of seating between source and receiver positions. Additional examinations of comb-filtering in stage enclosures were completed by Halmrast in 2000 [20]. It should be assumed that all these issues only increase in complexity when entering the realm of contemporary music. Due to constraints, such parameters were not included in the scope of this research, but they should be considered when making comparisons between classical and contemporary stage acoustics.

In a different arena, Ternström has recently utilized the in-situ measurement technique to address stage acoustic conditions in an opera chorus and a chamber choir [59,60]. His focus in this research was on the Self-to-Others Ratio (SOR), or the difference in SPL between the self-produced sound reaching the ear and the direct sound from other performers. Of course, because of the additional bone-conduction path created by singing, this SOR may be significantly higher for singers than for instrumentalists, even in the presence of a full chorus. Ternström found SORs among opera singers ranging from +15 dB (sopranos) to +10 dB (tenors) and in chamber choirs ranging from 0 dB to +8 dB. Looking back at the Gade data, however, it becomes clear that due to the great distances between members of the orchestra, it is possible to have even higher SORs between individual musicians (+19 to +23 dB ratios were used in Gade's large ensemble experiments [20]), which may have important relevance for contemporary works for small ensembles that include spatial instructions (*Atlas Eclipticalis* by John Cage, for example, requires musicians to be spread throughout the hall, creating wide separations between individuals) [46]. Inferring from Gade's construction of the EEL parameter, it can be said that SOR also is a possible contributor to subjective feelings of "Hearing Each Other" in an ensemble.

2.2 Psychoacoustics and Communication Acoustics

Two significant works were consulted in this research, and they will be described below. *Ecological Psychoacoustics* is a work edited by Neuhoff addressing the historical problem of applying laboratory psychoacoustics to the real world [44]. Cusack and

Carlyon examine the phenomena of Auditory Scene Analysis, or how we organize perceptually what we hear [8,14]. Auditory streaming is one type of perceptual organization, where we segregate sounds into groups based on selective attention to different parameters in order to place them in separate "streams" or auditory objects. Timbres, pitches, temporal order, rhythm, and spatial distribution of auditory streams can all change depending on the way sounds are grouped together. Parameters affecting the contents of auditory streams include harmonicity, amplitude variation, location of sources, frequency range of energy, pitch range of complex sounds, timbre, and overall intensity levels. Although these phenomena are easily observed under laboratory conditions, the complexity of auditory scenes increases in the real world, which implies that other techniques may be used to segregate sounds at a higher level. There is evidence that internal templates or schemas developed through prior experience are used for this purpose, and these schemas may even be multimodal in nature [14].

Jones addresses the question of attention described above by examining the anticipatory phenomena, where auditory scene analysis in an environment is used to direct attention to specific elements in the future [31]. Phase and rate information that are used to focus the attention at specific times will allow the listener to segregate the sound, provided it is constant in these parameters. Any time these parameters shift, there is a recovery time before the listener can react to the new values. These two essays in particular provide ideal methods for understanding the auditory scene analysis of a classical musician. This musician can expect certain parameter values to remain constant throughout a performance and utilize these to clearly segregate between instruments and melodic cues. However, it is likely that in the case of indeterminate and improvised music, such cues may not be available. Attention is therefore tenuous and must constantly adapt, using multiple parameters to segregate auditory information. Even within the output of a single musician, when the other player wants or is required to respond to specific information (e.g., a certain class of timbre, rhythm, or phrase) this information must be segregated from the musician's additional content, to say nothing of the presence of other musicians. In this situation, it is likely that the ecological understanding of auditory scene analysis becomes more important, and the higher-order, possibly multi-modal, parameters take precedence.

Valkenburg and Kubovy have proposed a modality-neutral theory of objecthood that takes into account these higher-level parameters [66]. This theory takes as its inspiration the synthesis of a traditional *Gestalt*[7] theory that is based on phenomenological principles, where perception is categorized by Husserl as *intentional* performance on the part of the Ego, or a form of "conscious perception" where information is perceived and grouped according to predefined and mutually exclusive formal principles [28], and Gibsonian theory, which emphasizes equally the pickup of information from all modalities and all parameters, later utilizing a combination of parameters to separate objects from this field of information. In their combined theory, the modality-neutral information picked up in a given environment is grouped initially based on Gibsonian principles and this pool of loosely-defined objects is susceptible to changing *Gestalt* segregation into "figure" (the object of conscious attention) and "ground" (the undifferentiated, unattended information). Neuhoff seconds this theory of interacting perceptual dimensions (within and across modes), and the fact that segregation may not be mutually exclusive.

One questionable conclusion from this essay is the choice of the term "object" over "event" to describe segregation. Cherry's "cocktail-party" effect (the ability to focus on one speaker among many others in a crowded room) [11] is put forth as one reason that "event" segregation would not be useful in perception (an event is segregated in time, while an object is segregated in space). The cocktail-party effect has long fascinated psychoacousticians, as evidenced by the large body of articles in this work's bibliography.

However, in contemporary music, due to the complexity of communication, perhaps it is necessary to consider the optimal situations for segregation of both auditory objects *and* events, for the reasons described above. Perhaps it is more akin to a different cocktail-party affect, where a listener divides his or her focus between the "object," or target (single speaker), and the "event," or background (entire room), constantly comparing to determine if he or she would rather engage in the current conversation or

[7] Psychological theory that assumes that the whole (the template, or interpretation of meaning) is more than the sum of individual elements (thoughts and experiences).

join another. The examination of such divided attention has been examined by Shinn-Cunningham and others, although their work often examines segregation in these situations based on the ability to separate information spatially, something which is not possible when multiple streams are coming from the same position, or a single stream is coming from multiple positions [9,10,11,15,33,34,35,53,56,57]. Rosenblum provides additional support for the "event" theory of perception. It has been shown through identification tasks of sound sources that mechanical, multi-modal properties and time-variant patterns of the event producing the sound have more correlation to the identification success-rate than do the objects created by segregating surface sounds. For example, the identification of the length of a dropped rod without visual verification is more correlated to the torsion and elasticity properties of the rod than to the frequency or intensity of its sound [54].

Communication Acoustics is a text edited by Blauert addressing the applications of psychoacoustics to different forms of auditory communication technology [4]. Communication is defined as *the transmission of a signal carrying information across space and time.* Blauert takes the traditional theories of auditory scene analysis and uses them to synthesize virtual auditory scenes. In particular, binaural processing, and interactive synthesis such as head-tracking contribute to authentic reproduction of auditory scenes. This is different from the theories described above in that the analysis used here is focused on a particular function, which is "communication." Not only must the listener receive and organize information; he or she also respond in kind. This may be as simple as Human-Machine Interaction, described by Möller, where the outcome is predetermined (buy a plane ticket from an airline) and specific data sets are required, or as complex as Human-Human Interaction, also described by Möller, such as a telephone conversation between friends or a telecommunication of some form between politicians, where the outcome is defined by the content and quality of the communication [43].

Kohlrausch et al. address audio-visual interactions, in particular the interactions that occur in the case of audio-visual mismatch (temporal-rate, and spatial disparities between auditory and visual information, as well as A-V asynchrony) [37]. In cases of temporal mismatch, the auditory information tends to dominate (the perceived visual information is integrated to the tempo of the auditory information), the visual information dominates (the perceived location of the auditory stimulus is shifted towards the actual location of the visual stimulus, known as the *ventriloquist effect*). In situations of asynchrony, listeners have a higher tolerance (integration ability) of auditory delays than visual delays. Other interactions also are observed in quality, distance, and motion judgments, indicating the importance of multi-modal perception. These findings are important in situations of *telepresence*, where the two modes of information are transmitted and reproduced separately, and any of the aforementioned conditions may occur due to degradation of the transmission. Fastl [17] and Möller [43], in separate essays, both address the issue of "quality" and how this relates to ease/efficiency of communication.

Jekosch [30] attempts to define sound quality and communication in terms of semiotic theory, a realm that can be loosely related to the work of Valkenburg and Kubovy [66] described above. Semiotics views sounds as "signs," conveying information about the source. Through this lens, communication can be defined as *an event that takes place between a perceiving human being and the world* and is *based on the processing of perceptual events as sign events*, as opposed to information and communication theory, which define communication as *an exchange of information between systems* and *the process by which spoken and written language are perceived as sign carriers*, respectively. The process of semiosis in this situation is described as matching the information up to a predefined *Gestalt* or schema, as described by Valkenburg and Kubovy. If the information does not match any particular schema, then the brain either "assimilates" the information by shifting it to fit a predetermined schema, or "accommodates" for the information by changing the schema itself to fit the information.

In relating the quality of a sound to communication, Jekosch presents three main relationships in a semiotic triangle: source, recipient and code relationships. For example, in the source relationship, the cues the sound creates about the source may be aesthetic, technical or functional. In the recipient relationship, the cues may be imperative, indicative, or suggestive of the recipient's response. The code (identification) cues may be indices (causality), icons (similarity) or symbols (arbitrary). In terms of complex musical communication, these relationships become important when determining which parts of a signal may be degraded or changed without affecting the communication. In other words, it is helpful to think in terms of semiotics when determining which parameters provide the aforementioned relationship cues and when determining the limits of assimilation and accommodation in improvisatory situations.

2.3 Telepresence

2.3.1 Definitions

As mentioned before, telepresence is gaining popularity as a means for the performance of contemporary music. It allows musicians to interact over large distances and collaborate in ways that would be unavailable otherwise. Telepresence has been defined by Henderson as "a wide variety of possible applications and methodologies for providing remote access to environments and events" but the particular application used in this research may be defined as "highly accurate acoustical auralizations for the rendering of immersive telepresence environments, where an individual may be immersed in the remote space in a perceptually-accurate and technologically-transparent manner" [27, p. 1]. This system is also referred to by other names, all of which help to expand its definition. For example, Cooperstock refers to such a system as "shared reality" as opposed to "virtual reality" because rather than creating an alternative world, telepresence attempts to recreate an authentic likeness of the remote musician within the virtual space [13]. Chew et al. call telepresence "distributed immersive performance" [11], and Schroeder et al. use the phrase "networked immersive space" [55]; both Chew et al. and Schroeder et al. emphasize the term "collaboration" in their work. Additionally, Schroeder et al. define "presence" as "having the experience of being in a place other than the one in which you are physically present."

In his research, Cooperstock addresses concerns inherent to the transmission system, such as video angle and resolution, latency (for example, limited by the speed of light, the travel time between Montreal and Los Angeles is at least 32 ms, and with the additional network delays, a total ping time of 97 ms is recorded), compression artifacts, and the tone-coloring effects of echo cancellation software, among others. Cooperstock writes: "Music serves as one of the most demanding of tasks from the perspective of both sensory acuity and sensitivity to timing, and is therefore, of greatest interest for study."

2.3.2 Implementations: ViMiC

Braasch has pinpointed other possible sources of acoustic degradation in telepresence, such as packet loss, feedback, and loss of spatial information, among others, and has proposed a system for introducing virtual spatialization after transmission, called Virtual Microphone Control, or ViMiC [6,7]. This is the system that has been used in the current research, and it will be described in further detail in the experiment protocol later in this paper.

2.3.3 Applications of Telepresence

Chew et al. have conducted experiments in audio latency and the performance of classical music. In these experiments, two virtuoso pianists (with extensive experience performing together) in remote locations attempt to perform Poulenc's *Piano Sonata for 4 Hands* as the latency between these locations is varied between 0 and 150 ms. In faster, rhythmic sections of the piece, a delay greater than 50 ms made coordination difficult and even impossible, whereas in slower, more lyrical sections, a delay of greater than 75 ms had adverse effects on the musician's perceived musicality of the performance. Unfamiliarity with the music (when players switched parts) further decreased the delay tolerance. While in indeterminate and improvised music, even when musicians have never performed together (as is often the case) it is certainly possible to perform under degraded conditions without the piece falling apart as it might in classical music performance, Cooperstock is insightful in this regard when he says, "at least subconsciously, users are aware of degradations and added delay, and alter their interaction behavior accordingly" [13].

Additional research has been conducted by Schroeder et al. [55] into other forms of telematic collaboration. In their experiments, two subjects are required to complete a rubix-cube type puzzle in a virtual environment from remote locations. In the virtual environment, it was found that the partners completed the task with equal accuracy and time taken as in physical reality, and felt that they were able to contribute equally to the completion of the task. Subjects also felt a strong sense of both presence and co-presence during the task.

An additional possible transmission degradation has been addressed in research by Garau et al. [24]. This research examines the effects of interruptions in the transmission or immersion in the virtual environment on the feelings of presence and comfort. Although the situation tested involved simple forms of communication and immersion of only one subject, it was shown that subjects were able to adjust and consciously or subconsciously ignore the interruptions (some were not even able to recall how many interruptions had occurred) and the sense of presence diminished slightly with each interruption but was not eliminated completely.

Although the research described above is important to the understanding and navigation of telepresence systems, it does not address the same type of situations created by contemporary music [47,48]. Pauline Oliveros is one example of a contemporary musician who uses the telepresence environment frequently for indeterminate and improvised music. Over the past 30 years, she has developed an "expanded instrument system" (EIS) for real-time processing of live instruments. Currently her work involves telematic performances between RPI, Stanford, and UCSD, among others. The performance protocol for these situations often involves attempts to eliminate the transmission degradations listed above, and relies on experimentation and classical acoustic principles to create the virtual environment. However, the flexibility of virtual reality and telepresence allows the synthesis of auditory scenes not found in physical reality. Therefore, it is possible that the combination of contemporary music performance and contemporary venue create a whole different set of acoustic requirements that could be used to further improve the quality of these performances. For this reason, the current research has been conducted within the telepresence environment.

3. Experiment Design

3.1 The Survey

Prior to embarking on this research, based on practical experience with contemporary music in a variety of venues, a survey was developed to determine the most important parameters for this research. The 21 parameters were divided into three groups. Four open-ended questions preceded the parameter list and are shown in Table 6. This survey was completed by nine musicians.

Table 6: Introductory Survey

1.	What is your instrument?
2.	Please briefly summarize your experience with improvisation and what goals you aim for when improvising or performing indeterminate music.
3.	Do you think communication is more difficult in degraded conditions when improvising or when performing written music (no conductor)? Why?
4.	Please describe your ideal ensemble and ideal venue for improvising or performing indeterminate music.

Although there were only nine responses, many musicians attested to playing more than one instrument. The overall distribution of instrument types is shown in Figure 1.

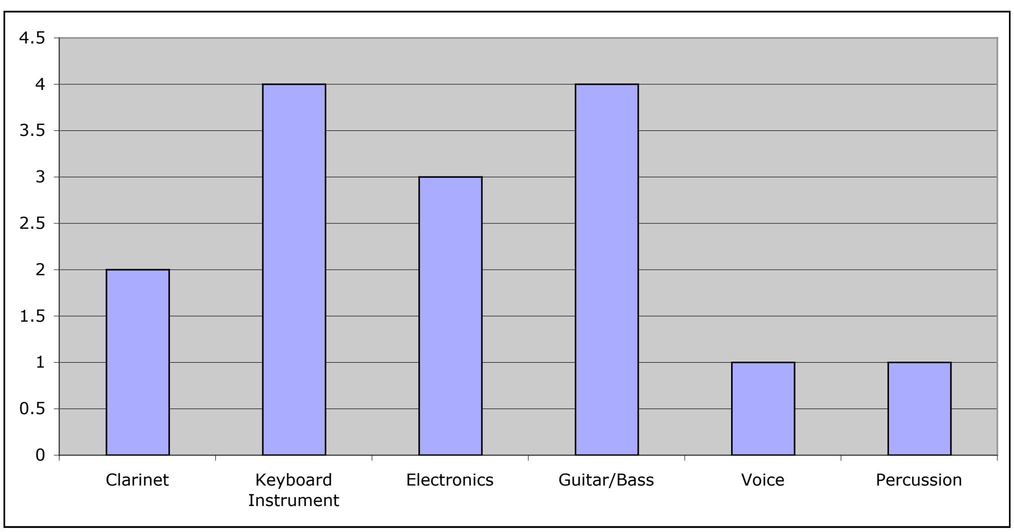


Figure 1: Instrumentation of Survey Subjects

The level of experience between subjects was varied, but all had some experience with performing improvised music, and the majority also had experience performing and composing indeterminate scores. The distribution (number of subjects out of 9) and average level of experience in each area (out of 5, with 5 being "very experienced") is shown in Figure 2.

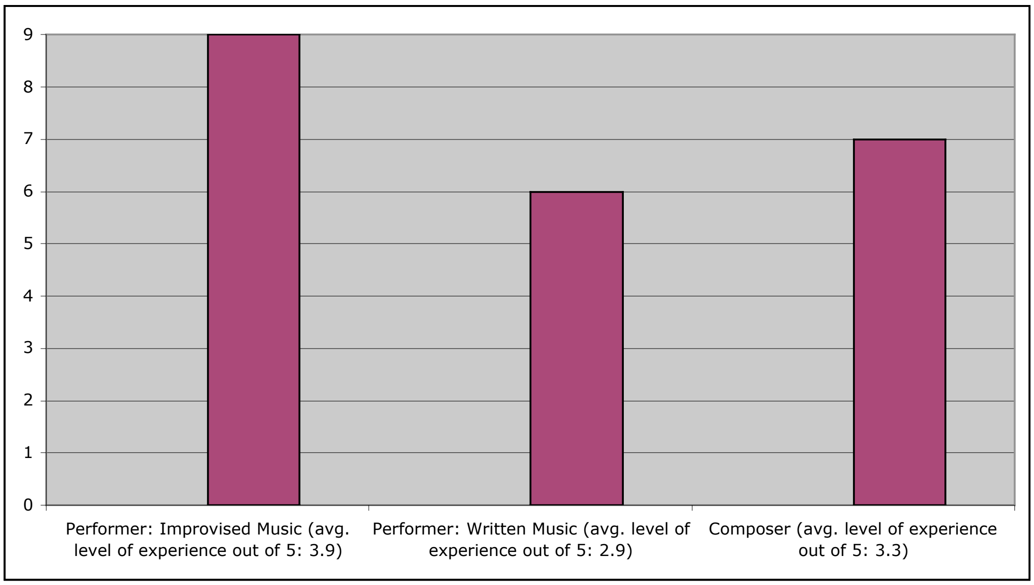


Figure 2: Experience Level of Survey Subjects

The responses to the second part of question 2 ("what are your goals when performing?") were complex and diverse, but the majority of subjects indicated the importance of communication clarity and good listening strategy. Some of the responses are listed in Table 7.

Table 7: Sample Subject Responses to Survey Question 2

"Follow the instructions of the score with clarity and subtlety."
"Defined, precise interactions with the ensemble."
"The performer's thought processes become visible to the audience."
"Create a meaningful piece of music by building upon, reacting to, and communicating w/the other musicians."
"Bring together diverse approaches and offer the possibility of unexpected musical intersections and encounters."
"In realizing indeterminate music or structured improvisation I try to best approach the intention [of the composer]. With free improvisation, I try to listen and find the best compliment to the situation if I can, or do something completely different."
"The main thing that differentiates 'improvisation' from 'composed' music is that improvisation generally implies a sense of shaping the piece to the environment of the space - and in a performance that is the hall, the other performers and the audience."

In response to the third question (Is communication more difficult in degraded conditions when improvising or when performing written music?), all but two of the subjects felt that communication was negatively affected by degraded conditions. The subjects who disagreed offered the following comments:

> 1. Not at all. Possibilities always exist for improvement.
>
> 2. I think it depends on the piece. "Improvising" can be also a set mobile of fragments that fit together in many different ways. I often question whether "improvisation" actually exists or is just a collection of previously learned ideas. Written music can get in the way when the performers are so concentrated on what they are "supposed" to play that they lose track of the moment, and stop listening and actively participating. But to briefly answer your question: no, I don't think it should matter, but the results will be different so you have to predict what will happen. I predict with a conductor it will be more hierarchical...you have the AMM-style ‘improvisation’ where there was a specific intention NOT to listen to each other, with quite striking results.

Although there seemed to be some confusion with the wording of this question, 6 of the remaining subjects seemed to feel that improvisation was more difficult in degraded conditions, while one subject felt strongly that both styles of performance (improvised and composed) were equally difficult in degraded conditions. Some of their responses are shown in Table 8.

Table 8: Subject Responses to Survey Question 3 (Is communication more difficult in degraded conditions when improvising or when performing written music?)

"Communication is difficult when the levels are too loud, monitoring is bad, and when performers are not facing each other."
“A conductor helps to keep a tight leash on things.”
“Regardless of the style or purpose of music being performed, it is crucial to be able to communicate w/ other members of an ensemble, either visually or via clear audibility of each musician. Improvisation is not ‘easy’ - but it is much harder if I have to guess at what other members of the ensemble are doing.”
“I'd think this would be pretty obviously a greater problem when improvising. Interactive decision-making requires good communication.”
“I think that clear and accurate perception of all players' contributions is essential for improvisation. This can be achieved through either good acoustics or through accurate monitors. But I believe that successful musical communication absolutely requires high-quality listening conditions.”
“Assuming ensemble conditions, I would surmise that degradation is an equal hindrance to communication in both settings. For written music, the score prescribes relationships that I must adhere to. These relationships invariably depend on the ensemble, so I have to listen carefully. In improvised music, I must listen carefully to the ensemble and place my improvisations accordingly.”

In response to the fourth topic, the list in Table 9 indicates just a few of the criteria listed by the subjects for an ideal ensemble.

Table 9: Subject Responses to Topic 4 (Ideal Ensemble)

Small Ensemble (2-4 players)
Familiarity or similarity between musicians
Instruments with good timbral and registral balance (similar or complementary)
Musicians who are sensitive listeners, don't feel compelled to play constantly (improv), think critically and don't overcomplicate situations (indeterminate music)
General virtuosity, dexterity, musicality, exploration, and inspiration

And for an ideal venue (see Table 10):

Table 10: Subject Responses to Topic 4 (Ideal Venue)

Small Venue
Good/supportive acoustics (clarity, balance and blend; reverberation time least important)
Dynamic range, audibility (able to hear self and others)
Control over proximity/spatial arrangement between musicians
Quiet venue (not a bar)
No need for further amplification to fill the room
Presence/interaction of Audience
No preference

The survey then listed 21 parameters, separated into three categories, shown in Table 11. The subjects were asked to rank the parameters within each category and then rate them individually on a global 1-10 scale, with 10 being the most important, in terms of communication during performance.

Table 11: Survey Parameters

	ENSEMBLE:
1	Number of musicians
2	Instrumentation (balance in register, timbre, directivity)
3	Familiarity and previous experience with musicians
4	Eye contact with other performers
5	Full visual clarity of other musicians
6	Audience (distance, size, placement)
7	Distance between musicians
8	Ability to distinguish between individual voices (audio only)
	GEOMETRY:
1	Size of space
2	Reverberance of space
3	Ratio of volume between hearing yourself and hearing others
4	Amount of silence in space
5	Position in space (room with or without stage, center, balcony)
6	Common aural space between all musicians (telepresence situations only)
	COGNITION AND ADAPTATION:
1	Familiarity with space
2	Knowledge of surrounding environment
3	Clarity of visual feedback - practical communications (other musicians, audience)
4	Authenticity of aural architecture (telepresence situations only)
5	Purpose (live performance or recording, ability to edit and cut material later)
6	Tactile issues (feedback and influence)
7	Intended length of performance (if length is determined beforehand)

The results of the questionnaire showed some subject confusion with the categorical rankings, and so these responses were discarded. The global ranking, based on the mean values of the individual 1-10 ratings from all 9 subjects, is shown in Table 12.

Table 12: Subject Ranking of Survey Parameters (Boldface print indicates parameters chosen for the subjective tests)

Ranking	Global Parameters	Mean ratings (10 is highest)
1	**Ratio of volume between yourself and others**	9.44
2	Amount of silence in space	8.22
3	Familiarity and previous experience with musicians	8.00
4	Common aural space between all musicians (telepresence)	7.2
5	Instrumentation (balance in register, timbre, directivity)	6.56
6	**Reverberance of space**	6.56
7	Ability to distinguish between individual voices (audio only)	6.44
8	**Eye contact with other performers**	6.11
9	Number of musicians	5.89
10	**Position in space (room with or without stage, balcony)**	5.78
11	Intended length of performance (also includes whether length is determined beforehand)	5.78
12	**Clarity of visual feedback - practical communications (other musicians, audience)**	5.67
13	Purpose (live performance or recording, ability to edit and cut material later)	5.67
14	Distance between musicians	5.11
15	Knowledge of surrounding environment	5.11
16	Familiarity with space	5.00
17	Authenticity of aural architecture (telepresence situations only)	5
18	Size of space	4.89
19	Audience (distance, size, placement)	4.78
20	Full visual clarity of other musicians	4.67
21	Tactile issues (feedback and influence)	4.11

Because this survey was constructed and distributed prior to the acquisition of the literature by Gade, not all the parameters utilized in his research were listed in the questionnaire, so unfortunately a direct comparison cannot be made. However, Gade's list was taken into consideration when determining the parameters for test in this research [20].

Based on a combination of the Gade list and the results from this questionnaire, the 5 parameters (listed here with their respective provenance) shown in Table 13 were chosen.

Table 13: Parameters chosen for current research

1.	Self-to-Others Ratio	listed in Gade, Ternström and questionnaire
2.	Early Reflected Energy	listed in Gade, Ueno, and Dammerud
3.	Reverberant Energy	listed in Gade, Ueno, and questionnaire
4.	Sound Source Position	listed in Gade, Meyer and questionnaire
5.	Visual Connection	listed in Chew, Cooperstock, and questionnaire

The parameter of early reflected energy, although listed in the Gade text, was not included on the survey but was later determined as important due to the large influence Gade's early energy measurements (ST1 and EEL) have had on the development of stage acoustics. The other parameters tested by Gade, "Time Delay" and "Timbre," as well as several of the parameters rated highly in the questionnaire (directivity, background noise, commonality of space, number of players, masking of players, etc.) were not included in the tests due to time constraints and the limitations of the experimental setup, but they will be considered in future research. The lower half of the questionnaire (parameters 11-21), with the exception of #12 (can be grouped with #8), were eliminated or held static during the tests. The exact values chosen for these parameters will be listed in the experiment protocol later in this work.

3.2 Musical Score

The music chosen for these experiments was taken from a work by Christian Wolff from 1964, *For 1, 2 or 3 People* [68]. This work is an ideal example of indeterminate music, perhaps one of the seminal open notation works written in this genre (other composers who engaged in this style include Morton Feldman, Earle Brown, John Cage, Cornelius Cardew and Frederic Rzewski). Although the possibilities for stylistic variety under the umbrella of "indeterminacy" are vast, and the concept itself has gone back as far as Charles Ives, Wolff's particular brand of indeterminacy is unique in that it truly requires the musicians to listen to each other and the space in order to create musical *continuity*. Nyman writes [46, pp. 66]:

> While Cage was moving towards this kind of indeterminacy involving pre-performance determinations (rather like traditional composition, but with shifted emphasis), Christian Wolff was evolving an indeterminacy in which all the decisions were to be made *during* performance, not by providing sound material to be realized on the spot (like Feldman and Brown) but by creating a chain of unpredictable situations which would only be brought about *through* the act of performing.

Wolff noted that this style left "chance completely outside the performer's control by making his *ear* the vehicle" [46].

When determining the musical material for these experiments, many other composers with unique indeterminate styles were considered as excellent candidates. For example, Cornelius Cardew's work with the Scratch Orchestra began in 1969 and "offered radical alternatives to the conservatism of most other forms of contemporary music, which are still largely tied to narrative models of expressive rhetoric and linear continuity...its Utopian vision of open enquiry and unfettered exploration, of an all-inclusive form of social music-making and performance" [51]. However, a major influence in the Scratch Orchestra performances (although they occasionally performed music by Wolff) focused on both aural and visual aspects, as noted by Cardew: "The word music and its derivatives are here not understood to refer exclusively to sound and related phenomena (hearing, etc.)." Therefore, the Wolff piece was chosen because, while it is certainly influenced by the visual environment, it is not completely dependent on such a connection, and there is no particular interpretation of the score that would completely eliminate the acoustic sensitivity of the performance. In other words, an acoustic deficiency, while it could be assimilated into the performance, could not be ignored altogether and the performance would certainly be altered by this deficiency.

For 1, 2, or 3 People can be performed with, as indicated by the title, 1, 2 or 3 performers. The score includes 10 pages with a spatially unique set of events (~20 per page) that are meant to be separated between the players in any way, so long as no event is played by more than one person. Each page can be performed as a single piece or in tandem with other pages. Wolff includes 2 pages of detailed instructions with his work.

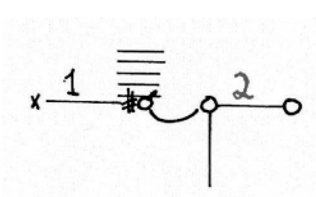
Nyman writes about the piece [46, pp. 67]:

> One player does not mover through these events regardless of what the other player is doing (as is the case in Cage's music). What one player does depends for its initiation precisely on what he hears the other performer playing [...] Each player makes his own particular continuity of structural units out of the common reservoirs of [material], and what he plays, and when, depends on which cue he hears, or perhaps fails to hear. But the rules are not to be followed at all costs: if both players are waiting for cues at the same time then, instead of remaining silent for ever, they have to work out a solution on the spot.

Through direct correspondence with Wolff and analysis of the score eight events were chosen to highlight the elements of the music felt by the author to be most acoustically sensitive. The events are shown and described below. The music is distributed evenly between the two musicians: two events are chosen for each condition, the musicians are given 30 seconds to play the events and then rate their ease of communication in each condition. The two musicians are never given the same events, in keeping with the rules of the composition, although the events are taken from different pages of the piece (four were from Movement I, four from Movement V).

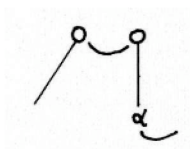
Open circles are not restricted in any way. Although they are referred to here as long tones, it is clear from the performances that they can be interpreted as entire phrase segments, depending on the acoustic situation.

3.2.1 Christian Wolff *For 1, 2, or 3 People*: Event 1



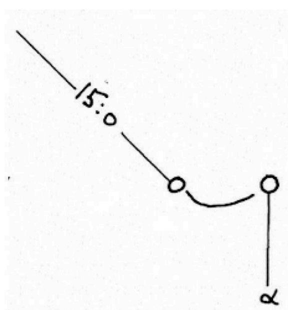
The musician is instructed to play anything (x) followed by one legato change to any aspect of the note (indicated by the one in black) and tied into a long tone of determinate pitch (F# in bass clef or D# in treble clef). The pitch should be bent up slightly and then a coordination between musicians is required (indicated by the tie). This coordination means that the pitched note should begin anytime, be held until another sound (from the other musician) begins, held still until that heard sound ends, and then the musician should move on. This movement includes two legato timbre changes (the red two) into a single long tone started and finished any time.

3.2.2 Christian Wolff *For 1, 2, or 3 People*: Event 2



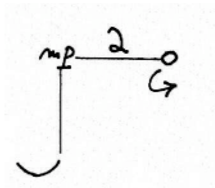
The diagonal line indicates a coordination between musicians. The note at the end of the line should be played directly after another sound in the environment (created by the other player). This note is open (any pitch, length, and timbre) and should be held until another sound is heard, then changed almost simultaneously with that sound (tie, straight line, and alpha symbol). This new note should be ended before the other sound finishes.

3.2.3 Christian Wolff *For 1, 2, or 3 People*: Event 3



The “15:0” along the diagonal line indicates that the musician should count to 15 (any tempo) before beginning the coordination (wait until another sound is heard before playing). Then the musician should play a single note, wait until another sound begins, change almost simultaneously with that note (straight line with alpha symbol), and hold until it stops.

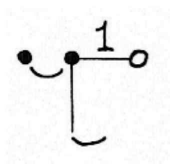
3.2.4 Christian Wolff *For 1, 2, or 3 People*: Event 4



The musician should wait until another sound has begun, enter while the note is still sounding (at a dynamic of mezzo-forte), hold until the sound stops, then move through two legato changes to some aspect of the note, and end on a long note that has some change in direction in space (the circular arrow: could be physical directivity change or melodic direction).

3.2.5 Christian Wolff *For 1, 2, or 3 People*: Event 5

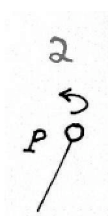
A short tone (less than one second, indicated by the black circle) should begin anytime,



be held until another sound begins, changed almost simultaneously, and finished before the other sound ends. The musician should then move through one legato change to any aspect of the note into a long tone that is started and finished at any time.

3.2.6 Christian Wolff *For 1, 2, or 3 People*: Event 6

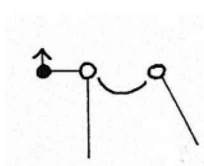
The musician waits until a sound is heard (from the other musician), then almost



immediately plays a long tone (piano dynamic) which is accompanied by a change in direction in space and two timbre changes (not legato), ended at any time.

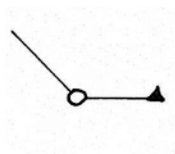
3.2.7 Christian Wolff *For 1, 2, or 3 People*: Event 7

The musician plays a short tone that is high in some aspect (the upward arrow), moves



via legato transition into a long tone began almost simultaneously with the other musician, and held until after the other musician stops. Then the note must end directly before another sound (in the environment, created by the other musician, or created by the individual).

3.2.8 Christian Wolff *For 1, 2, or 3 People*: Event 8



The musician should wait until a sound is heard in the environment, play directly after it, and hold a long tone any amount of time before moving on to a short event that is significantly dissonant from the other sounds in the environment (made by the self or others). This event can be finished at any time.

The transition between events can occur at any pace (usually this pace is dictated by the coordination markings. The events may be played in any order and last anywhere from one second to 30 seconds. The variety of responses is clear from the recorded content.

Depending on the distributed events, without any visual or highly sensitive transmitted sound, it is possible to have a standoff as both musicians are waiting for the other to create sound. This is crucial to finding the ideal acoustic conditions for such a performance. The open and darkened circles could be realized as phrases or single notes, and therefore the stylistic openness of the music allows for a wide variety of content, which can be clearly shown among the six subjects through a musical analysis, discussed later in this work.

A complete set of instructions is included in the Appendix.

3.3 Experiment Protocol

The equipment used in this experiment is listed and described below:

3.3.1 Physical Environment

The space used was the NYSTAR Virtual Acoustics and Telepresence Laboratory in the Gurley Precision Instruments Building at Rensselaer Polytechnic Institute (See the Appendix for photographs of the space). The space is acoustically isolated on a floating floor, with high-STC wall constructions and low RT values. The background-SPL levels with and without the telepresence equipment are listed in Table 14.

Table 14: Background Noise Levels in the NYSTAR Laboratory

	Equipment Off		Equipment On	
	Room A	Room B	Room A	Room B
dB-A	22	25	30	30
dB-C	50	40	52	48
dB(flat)	58	48	62	60

There are two main listening rooms, each supported by a separate control room. Room dimensions are 5.6 m x 4.9 m x 2.6 m for Room A and 5.9 m x 5 m x 2.6 m for Room B. In each room a Mitsubishi N623 LCD projector and a Da-Lite Cosmopolitan 6’x9’ projector screen (acoustically transparent up to approximately 8 kHz) have been installed. A circle of 8 Yamaha MSP5A self-powered monitor speakers (50Hz-40kHz) has been placed in a circle of 2.1m radius (the speakers that could not be placed at this distance due to room shape constraints have been processed with the appropriate delays and SPL corrections) at ear level (1 m height). The centers of each room (where the musicians are located) are separated (includes wall depth) by a distance of 4.9m and the distance from the center of each room to the projector screen is 1.7m. Each room is serviced by a computer running the Linux Fedora Core 6 operating system, connected via optical (Room A: RME Soundcard) and RCA (Room B) cables through an Alesis AI3 analog-digital interface. Control Room A also has a 3 GHz Mac Pro processor running Mac OS 10.4.11, which has been used in these experiments as a command center, receiving and sending Open Sound Control (OSC) over a UDP network to direct the experiments.

3.3.2 Portable Equipment: Installation and Calibration

In each room, a Sicura Tools Laser Level was used to properly position the speakers 45º apart and to position the microphones. A Behringer C2 Cardioid Condenser Microphone was placed at a height of 1m at each listening position. An Audix TR-40 Omnidirectional Microphone was placed at the halfway point between the listening position and the projector screen and was used to calibrate the transmitted sound fields in the two rooms. Each microphone was pre-amplified using a Soundcraft Spirit M8

Mixer (Room A) and a Mackie 1202-VLZ Pro Mixer (Room B). An Ono Sokki LA4240 Precision Integrating Sound Level Meter and a Duet portable audio interface (Apogee Electronics) with accompanying calibration software was used to calibrate the two rooms. The calibration included microphone calibration with a pistonphone at 94 dB, single-speaker and 8-speaker checks, as well as checks of the Internet transmission/ Virtual Microphone Control systems, which will be described below. The two rooms were set to unity gain using the volume controls on the mixers, software, and speakers. The Audix microphones were removed from the setup after calibration.

For the visual connection, a Sony 3CCD Megapixel DV-camera (Room B) and a Canon CMOS 3.1 mega-pixel HD-camcorder set to DV-quality (Room A) were positioned at the distance of the projector screen directly to the left of the screen, appropriately angled to capture a full-body image of the musician. This signal was transmitted via firewire to the Linux machines before being transmitted to the remote location and then projected over VGA connections onto the large projector screen. Photographs in the Appendix show the angle of the projected musician, and Figure 3 shows a diagram of the telepresence setup.

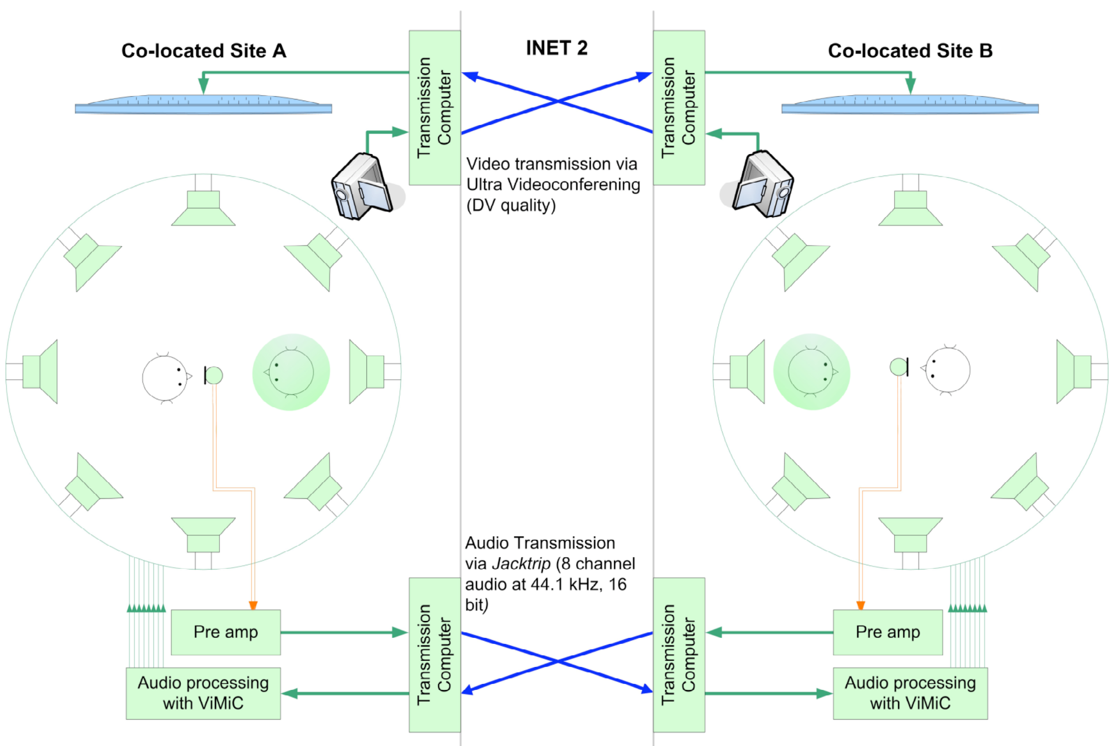


Figure 3: General Telepresence Setup [6]; Virtual Image of Musician in Green.

3.3.3 Audio Processing: Software

On the Linux machine in each room, the local, preamplified, monophonic close-microphone signal from the Behringer Cardioid Microphone was received and transmitted to the remote location using *Jacktrip*. *Jacktrip* is an internet transmission software for audio signals (one channel was transmitted from each room) developed at the CCRMA program at Stanford University.[8] Jacktrip utilizes an INET2 connection. The connection is made via the Jack server, and the overall transmission has a latency of 2.9 milliseconds for each machine (5.8 ms total), which is essentially undetectable by the human ear. Based on the speed of sound at room temperature of 343 m/s [2], the delay of the direct sound from one musician position to the other in physical space should correspond to a separation of 1.73 m, or 4.9 ms.

Since the transmission of audio-visual information over *Jacktrip* is limited at least by the speed of light (299,792,458 m/s), the transmission over this distance would only be in the nanoseconds, but the signal processing delay, which is caused by analog-to-digital conversion, data packaging, and routing processes, causes the latency to increase. These issues are all dependent on the performance quality of the hardware and software used for transmission. The current setup incurs a delay of 2.9 ms, which is very close to (and in fact, slightly improved from) the delay that would be experienced by musicians playing in the same location. The INET2 connection is capable of transmitting multiple Gigabits/s, but the current setup utilizes approximately 100 Megabits/s in both directions [5]. The basic requirements are 25 Megabits/s in one direction for DV-quality video, and less than 5 Megabits/s for 1 channel of high-quality audio in each direction. The system was calibrated at the listening position. The connection protocol (Linux Terminal) can be seen in the Appendix.

Following connection via the protocol listed above, the *Jack* server is routed internally in the remote computer to Pure Data (open-source software developed by Miller Puckette[9]) as shown in the *Jacktrip* signal routing diagram (see Figure 4).

---

[8] http://ccrma.stanford.edu/groups/soundwire/software/jacktrip/.

[9] http://puredata.info/.

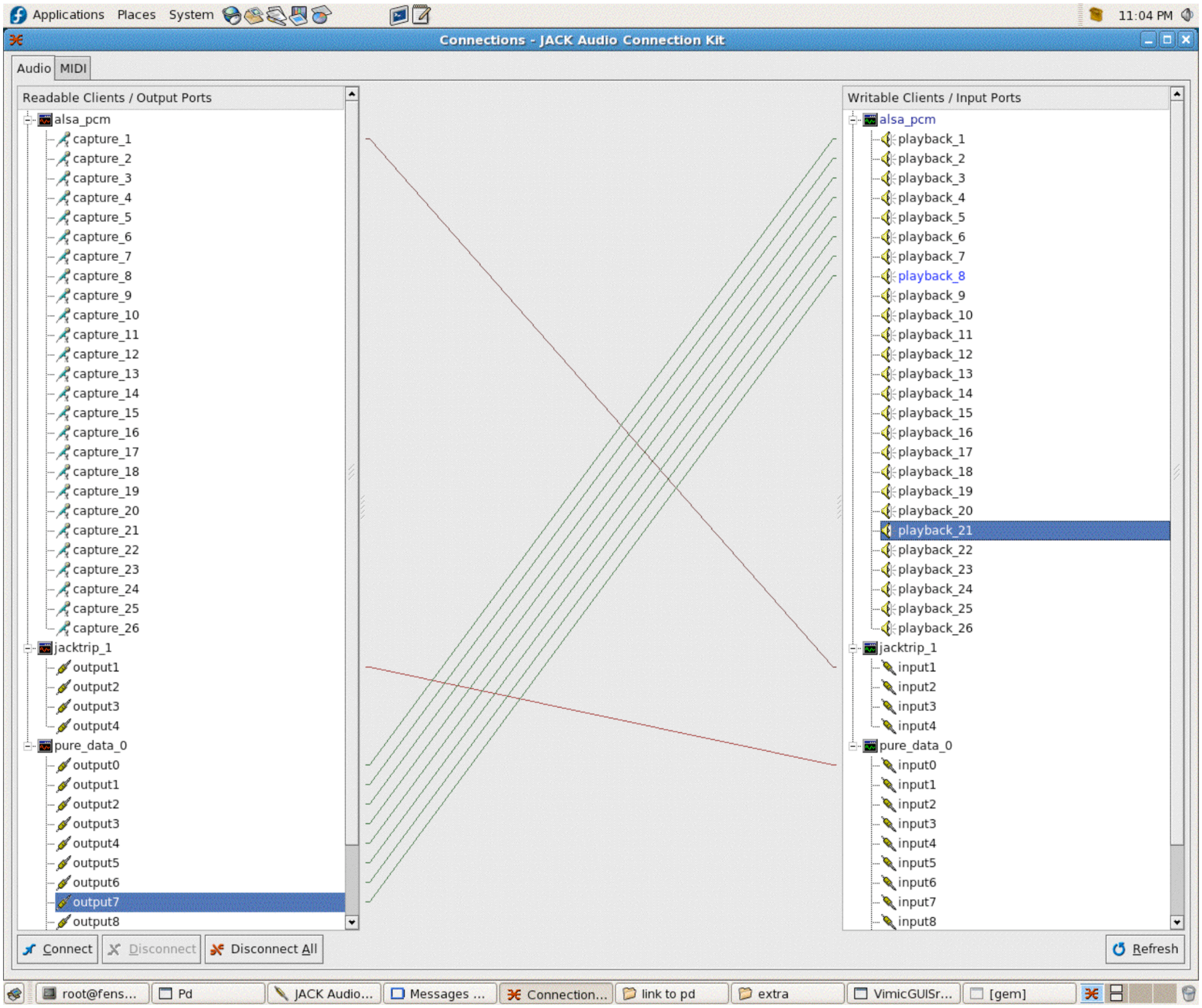


Figure 4: *Jacktrip* signal routing diagram

The video signal is transmitted over the McGill Ultra-Videoconferencing System (version uv-2.0.a0), software that has been designed by Cooperstock and others at McGill University.[10] The connection protocol can be seen in the Appendix.

---

[10] http://ultravideo.mcgill.edu

3.3.4 Audio Processing: ViMiC

When using telematic communication software such as *Jacktrip*, the transmission latency increases the audibility of echoes at much lower sound levels caused by the feedback loop between speakers and microphones at each location. Echo-cancellation systems, although they are useful for one-way communication (temporary suppression of one channel is highly effective), in bi-directional music communication, whole segments of the performance can be suppressed and the transmitted sound is often colored or distorted as a result. The most effective way to eliminate this problem (although there are still some issues with feedback) is to use a closely-positioned microphone; however, while this procedure alleviates some of the feedback, it eliminates the spatial information that would be captured by a room microphone array.

To deal with this issue, Virtual Microphone Control (ViMiC) has been developed to interact with Pure Data and other applications [6]. ViMiC works by assuming a virtual array of microphones, using the *Sound Field Renderer* external (algorithm developed in C++ to be used in Pure Data) to calculate the difference in gain and delay between the sound source (the close-microphone "anechoic" signal) and the virtual microphones (based on user-generated microphone positions, axes and directivities). The difference in gain and delay between virtual 1st-order reflections (using user-generated room dimensions and the Image-Source Method) and the virtual microphones is also calculated using the *Sound Field Renderer*, which is essentially a multi-tap delay unit, and the late reverberation is calculated using a multi-channel feedback-delay algorithm, a 16x16 Hadamard Matrix to achieve the appropriate reflection density, and IIR Filters (user-generated cutoff frequency) to simulate frequency-dependent surface absorption. The difference between ViMiC and a system like Vector-Based Amplitude-Panning (VBAP) is that VBAP relies on interchannel level differences whereas ViMiC depends on interchannel time and level differences. Figure 5 shows the basic flow diagram for this program, and a full list of controllable parameters is shown in the Appendix.

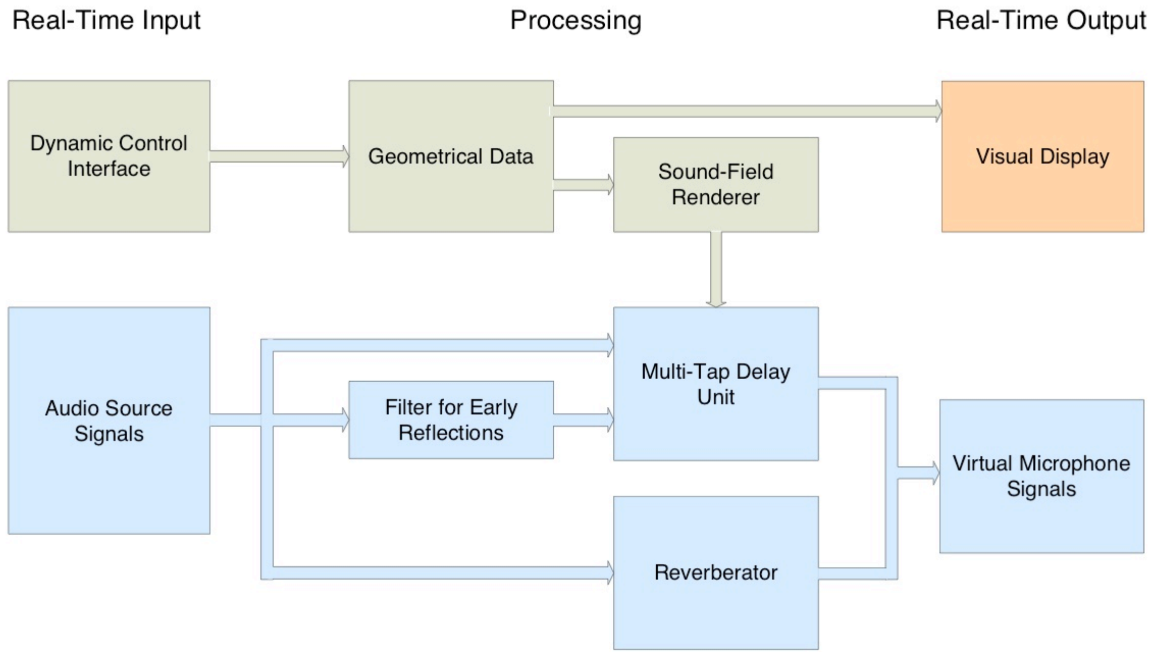


Figure 5: Flow diagram of ViMiC procedures [6]

In these experiments, 8 virtual microphones were used, one for each output speaker. The virtual source was positioned 1.73 m from the listener (corresponding to the measured distance between listener and projector screen), and the virtual microphones were positioned 0.11 m from the array center. The directivity used was a cardioid pattern, which is represented by the following equation [7]:

$$P = \frac{(1+\cos\theta)}{2}$$

where θ is the angle measured from the front of the array. The distance power used was "1," representing the inverse square law of –6dB/doubling of distance. A previously-developed patch was modified for the manipulation of specific parameters in ViMiC for this research (see Figure 6).

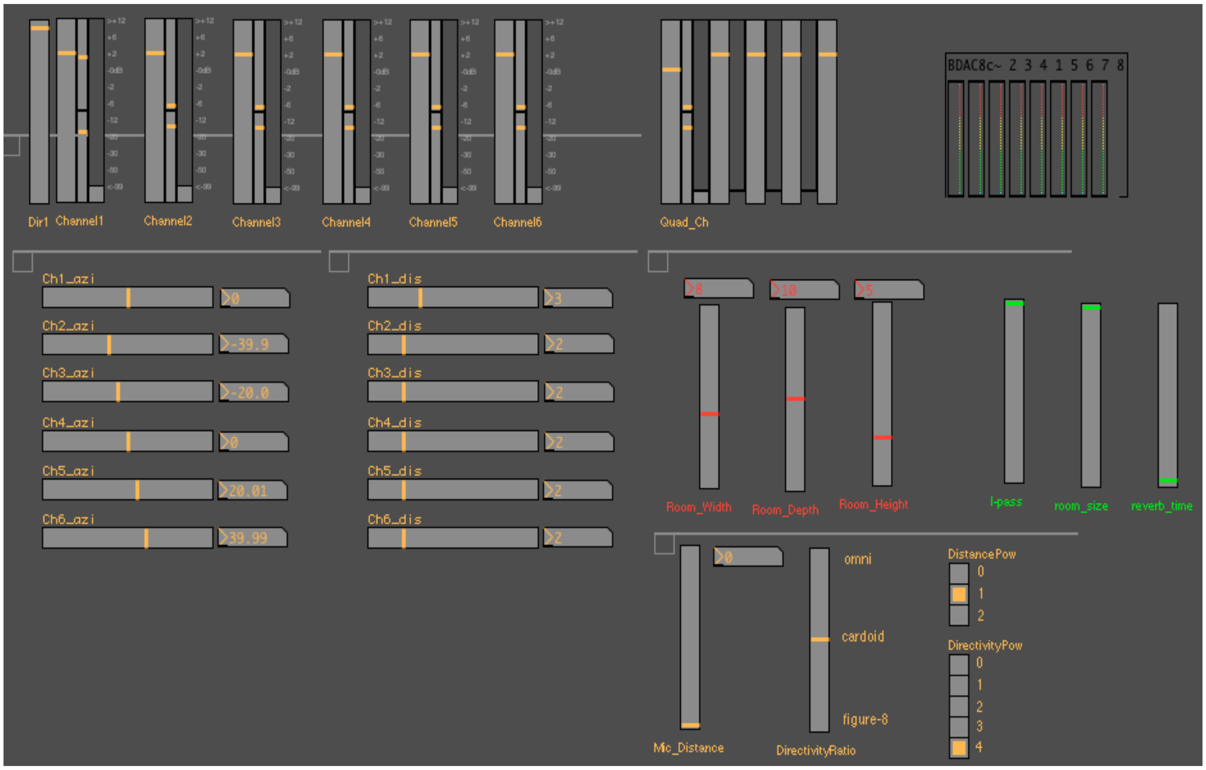


Figure 6: Sample ViMiC GUI

Typically, two parameters were varied in each test (a sample patch from one test can be seen in the Appendix), and the remaining parameter values were held static. These values were determined through the creation of a model in CATT Acoustic of a small gallery (10 m x 15 m x 5 m) typical of contemporary music performance. The CATT Model and resulting acoustical data are shown in Figure 7. The T_{30} shown in Figure 7 has a mid-frequency value of ~2.1 s. This was chosen for the extreme case of the Reverberant Energy test. For the rest of the tests, the value shown at 10kHz, about 0.45 s, was used to prevent excessive reverberant energy from interfering with the experiment. The cutoff values of 8 kHz (early part) and 6 kHz (late part) also reflect the curve of frequency-dependent decay times shown in the top portion of Figure 7.

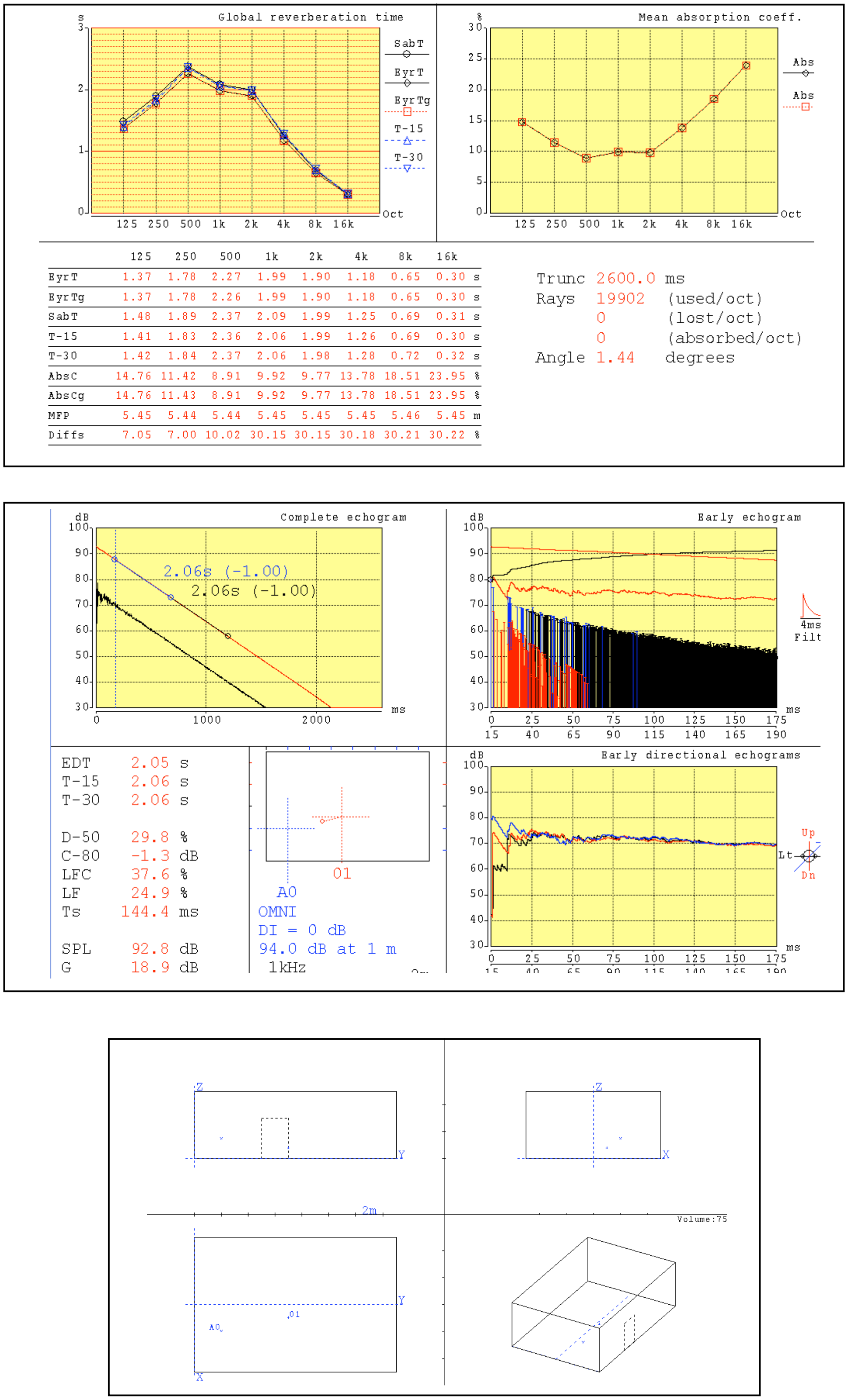


Figure 7: a) Frequency-dependent Reverberation Time (*T_{30}*) of CATT Model; b) Energy decay curve, showing early reflections, of CATT Model; c) 3D-Model of virtual room, CATT Acoustic

3.3.5 User Interface

The music and Graphical User Interface (GUI) were presented to the musicians via MacBook Pro laptops (6 subjects) and a G5 Macintosh processor (2 subjects) with Max/MSP 4.6 and Jitter 1.6. An image of the GUI is presented in Figure 8. Each GUI patch was wired to send OSC commands to a Mac Pro 3GHz processor in Control Room A. The image of this patch is shown in Figure 9. Each musician was presented with the same GUI. The musical material was presented on the upper-right corner along with a counter showing the number of the current trial (synchronized via the command computer), and the instructions were presented on the left-hand side of the screen. All user interactions were constrained to the red box on in the bottom right corner of the screen, and these interactions will be described below.

After both users activated the button labeled “Begin,” the two patches sent a “bang” (Max message indicating action) to the command computer over the network and this bang triggered the command patch to read the conditions for the first trial from multiple text files (generated in Matlab, described in Section 3.4) and sent these conditions to the respective computers (acoustic conditions were sent to Pure Data on the Linux Terminal, musical score and timing data were sent to the Max/MSP GUIs on the Macintosh Laptops). The “Begin” button also created 2 new text files on each laptop, which would be used to record the preference data and the subjective comments. The musicians had approximately 30 seconds (shown on the counter in the bottom-right corner of the box) to play 2 musical events and after this period, the set of radio buttons to the left of the box was made available for the musicians to rate the preceding condition and a random number was generated on the set of radio buttons in order to eliminate the influence of previous judgments on the current condition. After rating the condition and entering any additional comments into the text box on the right, users activated the button labeled “Click here to move on,” which recorded their preference rating and comments into the text files created at the beginning of the test. The activation also sent a “bang” message to the command computer, which triggered the next condition. After 35 (SOR test) or 36 (all other tests) trials, the command patch triggered a pause and musicians were allowed a break of indeterminate length (approximately 10 minutes), after which the y were allowed to continue using the procedures described above. After 70 (SOR test) or 72 trials, the command patch triggered another pause to signal the end of the test.

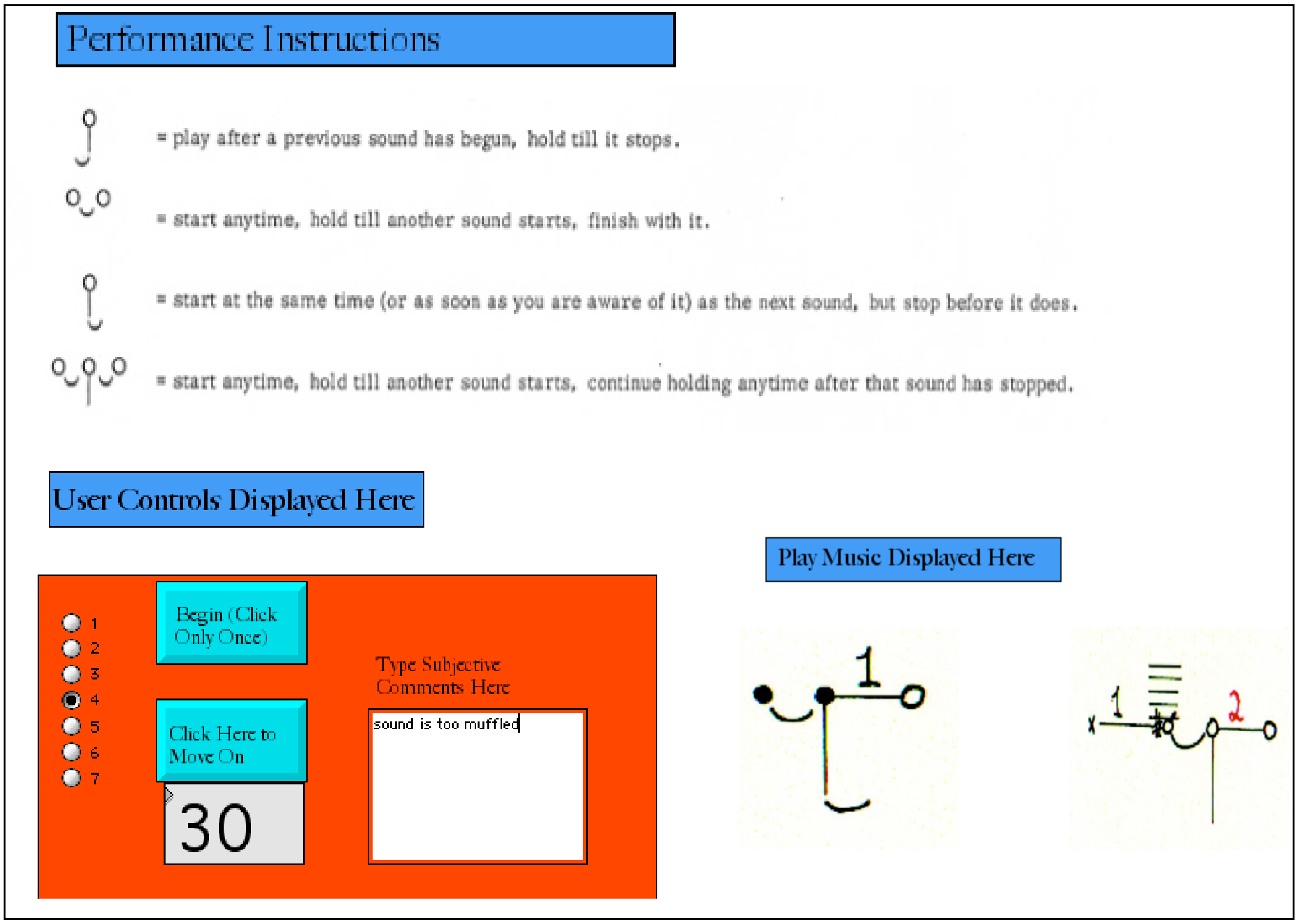


Figure 8: User Interface in Max/MSP 4.6

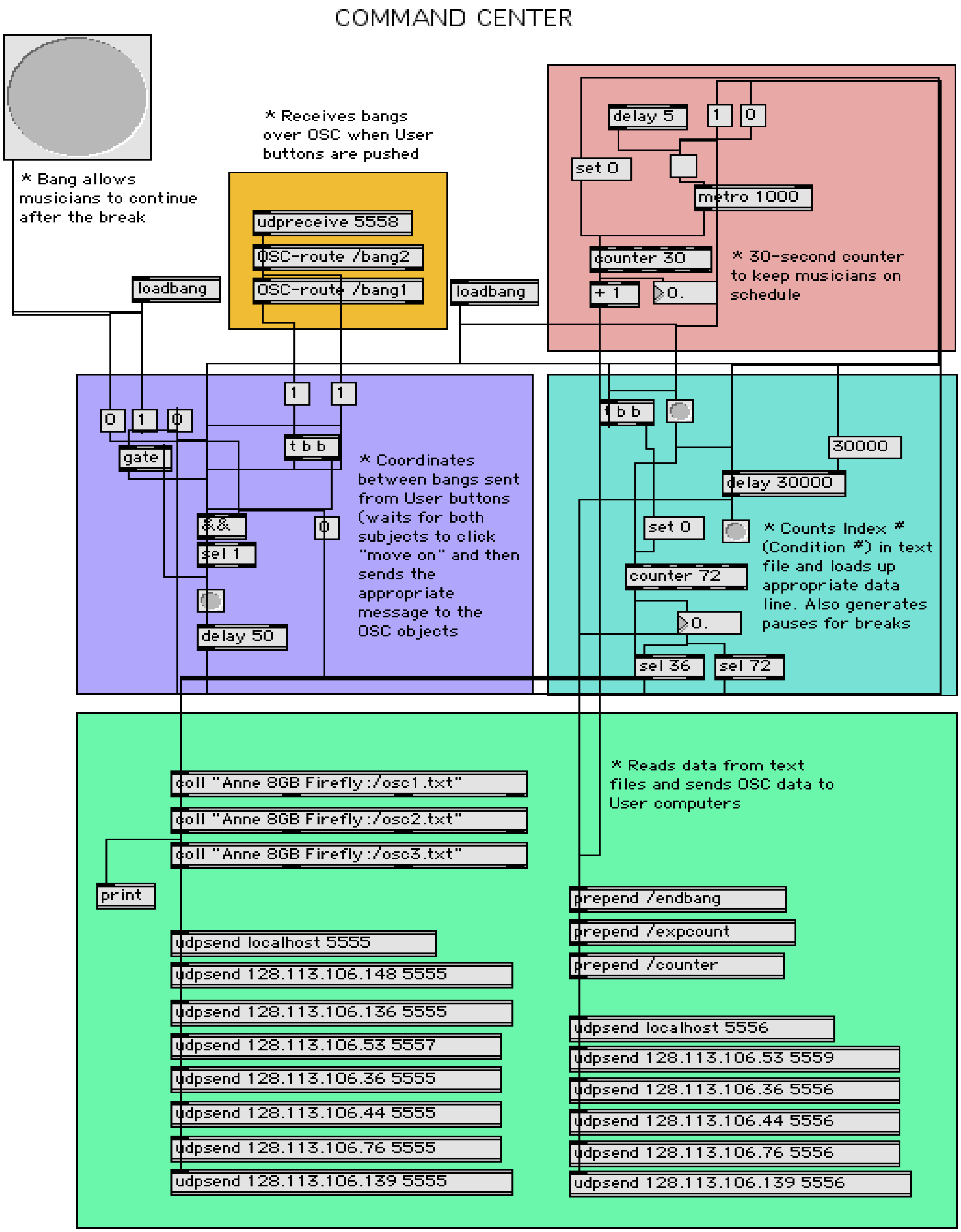


Figure 9: Command Patch (Max/MSP 4.6)

3.4 Experiment Design

3.4.1 Pilot Experiment

Initially, a pilot experiment was run using three subjects, in order to test the telepresence protocol and determine the best structure for the final experiment. Results from the pilot experiment were not recorded, and the main purpose of the setup was for discussion between the author and participants. Four parameters were tested: Transmitted Sound Level, Visual Connection, Network Latency, and Reverberation Time. In the Network Latency test, the 3 musicians attempted to synchronize to a rhythmic pattern. It was determined that although this experiment was interesting, it did not pertain to the musical material used in this research. The Visual Connection experiment was determined to have the strongest effect, slightly ahead of Transmitted Sound Level, and both of these parameters were chosen for the final experiment design. The Reverberation Time parameter, although subjects felt it did not significantly degrade or improve communication, had a strong effect on the interpretation of the score and the subsequent musical content. Therefore, it also was chosen for the final experiment design. Comparing the situations between Room A (1 musician) and Room B (2 musicians) it was determined that 1 musician in each room (2 musicians total) was sufficient to rate the acoustic parameters. Additionally, an entire movement from the Wolff score (~30 events) was loaded onto the screen for this experiment, and it was determined that this was far too much. 2 events per trial, chosen from a smaller pool (8 events), with 30 seconds of playing time per trial, would be sufficient. The subjects also preferred variation of a single parameter for each test, rather than a single test where all the parameters would be varied at once.

3.4.2 Final Experiment Design

Several possible test structures were available for this research. In the end, multi-dimensional scaling, a technique that has been used frequently in the past to determine the correlation between subjective impressions and objective parameters [25], was eliminated from this test because the goal was not to compare or rank different situations but to have an absolute rating for each individual situation. The 2-Alternative Forced Choice Test, a common psychoacoustic method, was also eliminated because the goal was not to determine if the changes in values were audible but rather how efficient and successful was the communication between musicians. A simple rating scale of 1-7 (7 indicating highest efficiency and preference) was chosen to accomplish this goal.

Because of the flexibility of ViMiC for isolating parameters perceptually (as opposed to natural environments, where multi-dimensional scaling would be necessary to extract the parameters and determine their influence) while keeping others static, each parameter was tested separately, with one exception. Because of the cross-modal influences and the ventriloquist effect (the brain's desire to integrate localizing information from auditory and visual perceptions), Sound Source Position and Visual Connection were tested together. In all tests, a Matlab script was used to create a randomly-ordered text file with all possible combinations of parameter values and musical events. In a given test, 72 conditions were created, in which the six parameter values were tested 12 times apiece to reduce the statistical significance of experimental errors. An entire test lasted one hour. As a result, each musician participated in the experiments for a total of four hours.

Based on the pilot experiment, the musical events were considered to have insignificant effects on the ratings (although effects on the performed content would obviously be present) so although all possible combinations were tested between the musical events and the acoustic parameter values, and each event was repeated several times, the combinations themselves were not repeated. The events were distributed according to the rules created by Wolff listed above, so that each musician received 2 events per condition, creating a combination of 4 different events each time. The order in which these events appeared, although which according to Wolff may be played in any order, was also randomized to eliminate a possible bias for the initial event. Each condition was presented for at least 30 seconds (although the counter was provided to keep the musicians informed of the time, it did not prevent them from taking more time if necessary) to eliminate excessive variation in the performed content that might add experimental error in the ratings.

3.4.3 Subjects

A total of 9 subjects participated in these experiments (3 of the subjects also participated in the pilot experiment), 8 males and 1 female. 7 subjects completed all 4 tests, and 2 subjects only completed 2 tests apiece, providing 8 complete data sets for each test. Subjects ranged between ages 23-47; the median age was 28. The average experience level was 20 years of training (all subjects had at least 10 years of musical experience). The subjects were paired with the same musician for all 4 tests (except for 1 subject who had 2 different partners throughout the tests). The subject pairings by instrument are shown in Table 15.

Table 15: Instrument pairings

Pair 1	Clarinet	Violin 1
Pair 2A	Soprano Saxophone (2 tests)	Dilruba (2 tests)
Pair 2B	Soprano Saxophone (2 tests)	Bass Guitar 1 (2 tests)
Pair 3	Percussion	Oboe
Pair 4	Bass Guitar 2	Violin 2

The overall instrumentation, therefore, included 3 winds, 5 strings, and 1 percussion. The players all claimed to have normal vision, and reported no hearing loss. The subjects' overall experience with indeterminate/improvised music is shown in Figure 10, and their experience with acoustics is shown in Figure 11.

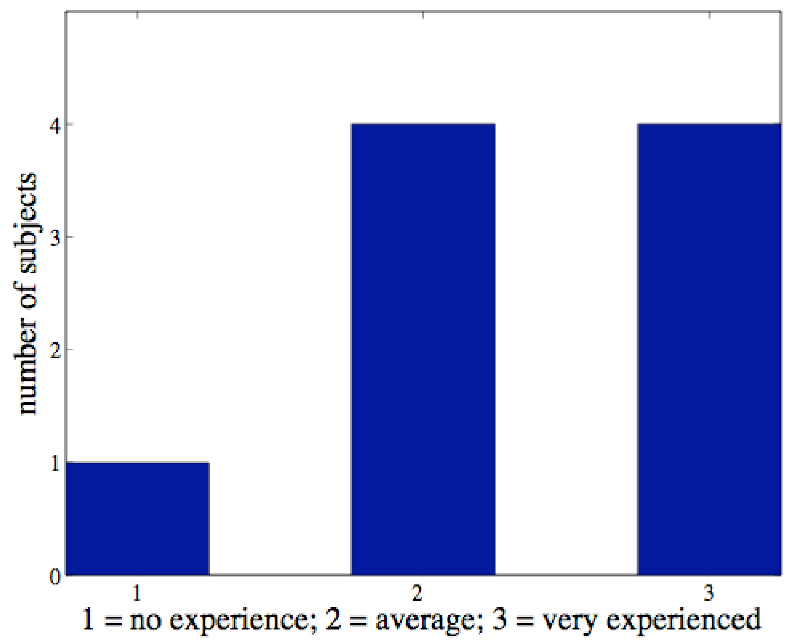


Figure 10: Subject experience with indeterminate or improvised music

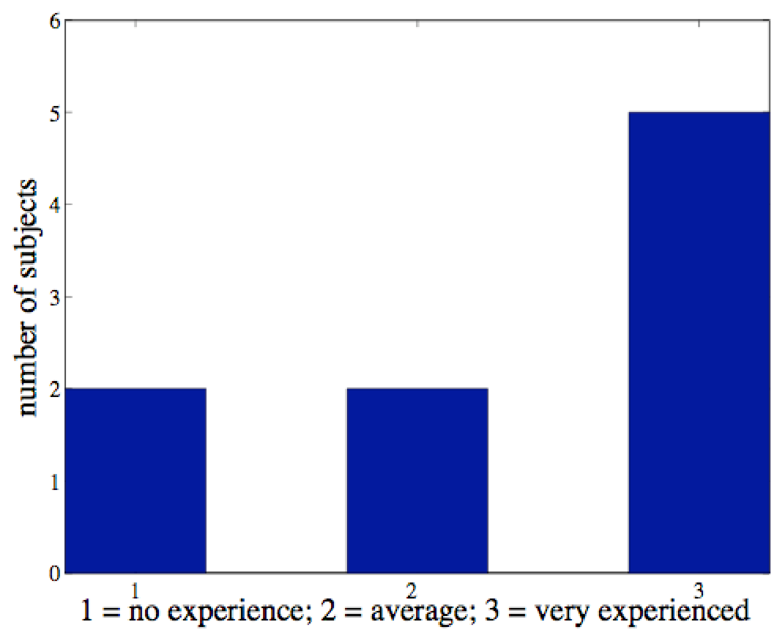


Figure 11: Subject technical experience with acoustics

4. Results

4.1 Test 1: Self-to-Others Ratio

4.1.1 Experiment Design

In this test, the calibrated transmission SPL (at unity gain) was labeled as a Self-to-Others Ratio (SOR) of 0 dB. The level of the transmitted sound varied between 7 different values and each combination of values was paired with a different set of Wolff events (10 possibilities) to create 70 trials. The values with relation to unity gain and the corresponding SOR are listed in Table 16.

Table 16: ViMiC Direct Sound: Acoustic Parameter Values

Transmitted SPL in dB	Self-to-Others Ratio in dB
+18 dB	–18 dB
+12 dB	–12 dB
+6 dB	–6 dB
0 dB	0 dB
–6 dB	+6 dB
–12 dB	+12 dB
–18 dB	+18 dB

The remaining parameter values are listed in Table 17.

Table 17: ViMiC acoustic control parameter values

Room Dimensions	10 m x 15 m x 5 m
Room Size	750 m^3
Source Position	0º (front of head in azimuth plane)
Source Distance	1.73 m
Cutoff Frequency (lowpass FIR filter)	8 kHz (early) 6 kHz (late)
Reverberation Time	10% Feedback (~0.45 s) avg. over all frequencies
Early Reflected Energy	0 dB (normal spherical attenuation)
Reverberant Energy	+4 dB (above normal spherical attenuation); 0 dB (Tests #3 and #4)

Due to some difficulties and deficiencies inherent in the setup, feedback and packet loss were introduced with the increase of transmitted SPL, so the absolute SPL values listed in Table 16 may be overestimated slightly (to save the subject's ears, the mixer had to be turned down for the higher-level conditions in some tests with louder instruments such as oboe and saxophone by up to 10 dB). However, as will be shown in the results, this variability did not significantly affect the subject preference ratings.

4.1.2 Results

Due to technical difficulties, only half of the data was recorded for two subjects (both violinists), resulting in 7 full data sets of 70 trials each. The means and standard deviations of preference ratings (calculated using JMP software) for each SOR are shown in Figure 12. The SOR values listed along the abscissa are ratios of self-produced volume in dB over transmitted level in dB (see Table 16 for relationships). The preference ratings along the ordinate axis are the ratings given by each subject after each trial, with 7 indicating highest ease/preference of communication, and 1 indicating lowest ease/preference of communication.

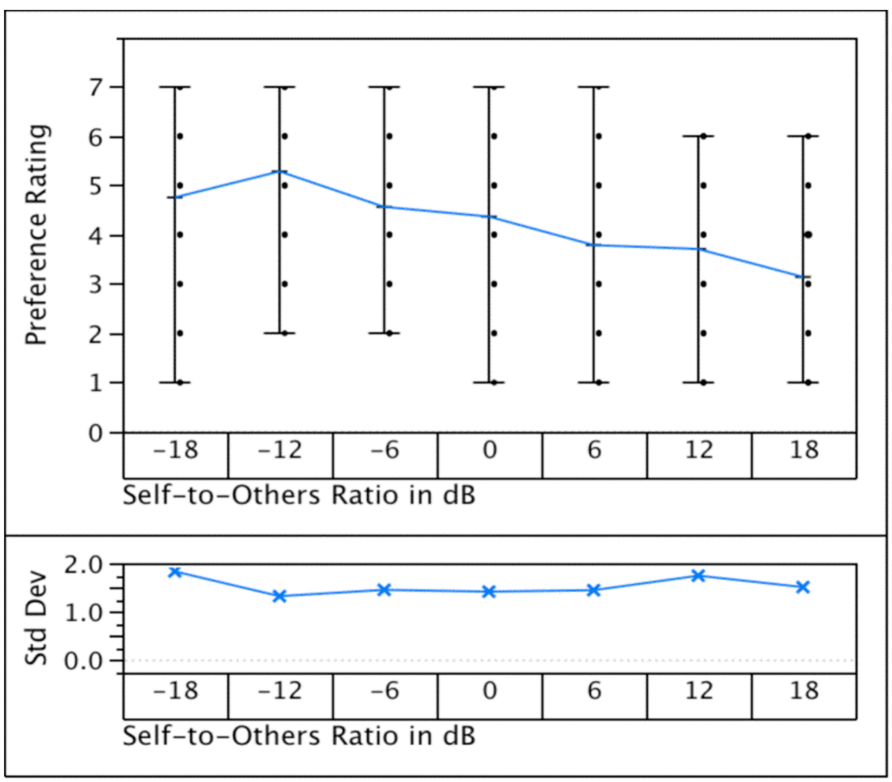


Figure 12: Mean Preference Ratings (with Range Bars) for Self-to-Others Ratio

In spite of the wide variability (data exists for all preference ratings in a few conditions) shown by the error bars in Figure 12, a trend is visible. Preference appears to increase in a nearly linear fashion with an increase in transmitted SPL, or a decrease in Self-to-Others Ratio. The highest mean preference is at an SOR of –12 dB (the remote musician is 12 dB higher at the virtual source position than the self-produced level, although this level was slightly attenuated following the inverse square law as it approached the listening position). Perhaps because of the increase in feedback and packet loss, the preference begins to decrease at SORs of –18 dB. It is possible that the excessive masking of the self-produced sound at this level could cross a masking threshold.[11]

In the literature, SORs between +23 dB and +10 dB have been tested. Traditionally, this parameter is limited by attenuation due to spherical spreading(the self-produced sound, having shorter air-conduction paths and the added volume from bone-conducted paths in some instruments, tends to be louder than the direct sound from other instruments in the physical environment). However, in telepresence, the possibility arises for negative SORs, and it is shown in this graph that this type of ratio is preferred. Of course, it is possible that the bone-conduction paths for some instruments, as well as the specific directivity patterns of the self and the remote musician, may cause these ratios to

[11] Shinn-Cunningham's work in spatially-separated, divided-attention, speech-masking tasks shows subjects crossing the 75%-correct mark at about –15 dB target-to-masker ratios and beginning to level off around +10 dB (the tests do not extend above this ratio). This could indicate a similar threshold as the conditions of her informational masking tests (if not the material) are similar to this research [57]. Kidd has examined the different non-speech signals that would cause energetic (peripheral) vs. informational (central) masking, and it seems safe to assume that the masking created by this research would be mainly informational [33]. Brungart also found that in situations where the signals were already highly discernible, the separation in distance would have little effect, as seems to be the case in this research [10].

be inaccurate indicators of the real difference at the ears between the 2 instruments,[12] but it is safe to say that the errors would be less than 12 dB, so the ratio preferred in this graph would still be negative. In order to determine if these instrument-dependent effects may have skewed the data, and also to compare with Gade's and Ueno's stage acoustics research, which showed significant differences between winds and strings, the data has been pooled by instrument groups and reanalyzed in Figure 13.

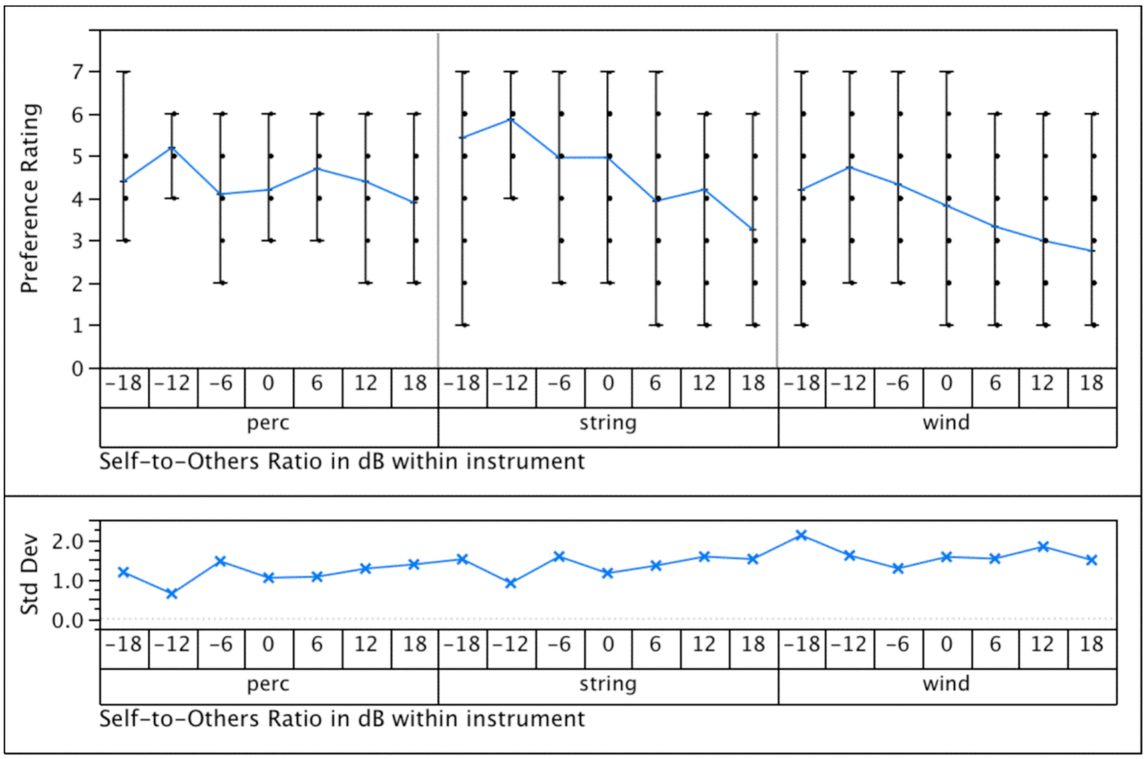


Figure 13: Mean Preference Ratings for SOR Pooled by Instrument Group

As the data show, in spite of some minor variations in the percussion data (only 1 subject played percussion, so the data set is small), the contour is the same across all instrument groups, and all three groups indicated SOR –12 dB as the easiest condition for communication.

---

[12] Ternström also shows an increase of about 3 dB when the signal is well correlated, or the same distance from each ear (the self-produced signal may fit this mold for wind players) [60].

However, before it is possible to assume any relevance or ideal parameter values from this data, it is important to determine if the data is statistically significant [26]. Because there is a chance that these trends are simply due to random distribution of data, it is necessary to analyze the variance, which can be done in several ways. For normal distributions, it is common to use a t-test for this, or an ANOVA test. However, the data in this experiment is most likely of a non-normal distribution, mainly because the data set is relatively small, and the variance is non-homogeneous. Therefore, a non-parametric test should be used. The Kolmogorov-Smirnov (KS) "goodness of fit" test was chosen because it makes no assumptions about the data distribution, it is good for data sets of unequal size and small samples, and it is sensitive to changes in both shape and location of distributions. The equation for the Kolmogorov-Smirnov statistic is given below:

$$D = \max_{1 \le i \le N}\left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i) \right)$$

where Y is the data set and F is the theoretical cumulative distribution of the data [38]. Using the "kstest2" function in Matlab allows the user to enter 2 data sets and compare them to determine whether these 2 sets are from different populations, or statistically significant. The "null hypothesis" assumes that 2 samples are from the same population, and therefore the difference between them is not significant. Determining whether or not the null hypothesis can be rejected is the goal of the KS test and the Matlab command returns 2 variables: H and P. H is either true or false (1 or 0); a 1 indicates that the null hypothesis can be rejected and a 0 indicates that it cannot. P is the probability (out of 1) that the data is from the same population, or that the null hypothesis is true. Therefore, if P is less than 0.05 (5%), H should equal 1, showing that the difference between the samples is statistically significant.

The KS test was applied to pairs of data (including all subjects) separated by SOR value. The P values for relationships between SORs across all subjects are shown in Table 18.

Table 18: Probability *P* from KS test (statistically significant relationships marked in boldface print)

Kolmogorov-Smirnov 2-way significance test (Probability *P* is significant if $P < 0.05$)						
12	6	0	–6	–12	–18	SOR in dB
0.14	0.096	**0.0012**	**4.57E-04**	**7.57E-09**	**1.61E-05**	18
	0.7939	0.1087	0.0656	**5.66E-04**	0.0768	12
		0.2061	0.2001	**1.11E-05**	**0.0039**	6
			0.8757	**0.002**	0.0855	0
				0.0351	0.6575	–6
					0.4429	–12

The boldface values in Table 18 indicate relationships that are statistically significant. SORs of +18 dB are statistically different from SORs of 0 dB to –18 dB, SORs of +6 dB are statistically different from SORs of –12 dB to –18 dB, and all SORs except for –18 dB are statistically different from SORs of –12 dB. This seems to indicate that –12 dB has crossed some sort of threshold below which communication is difficult, and that at this threshold, communication is easier.

Although the preference appears to drop above this, it cannot be claimed that an SOR of –18 dB is truly less preferred than one of –12 dB. It would be useful to test higher transmitted sound levels to determine the upper limit of communication ease, but this would cause problems with feedback and would likely breach the OSHA (Occupational Standards and Health Administration) limits for noise exposure. The plausibility of creating such tests without causing feedback will be examined in future work.

It can also be said from this data that a ratio of +18 dB is least preferred (communication is most difficult). On the other hand, one subject commented when performing at this level that, “sometimes communication is not required,” indicating that it is always possible to overcome obstacles, and that certain situations may call for these exact obstacles. At least now, however, it is possible to know what value of SOR would create such an obstacle.

To determine the instrument-dependence of this data, the KS test was performed between the groups and in spite of the similar contours, the winds were found to be statistically different from percussion and strings, although percussion and strings were found to be from similar populations. KS tests were performed between SOR pairs within each instrument group, and winds and strings were both found to have similar *P* values, with less significance on the edges. Percussion was not found to have significant data in *any* pair relationship, but since there was only one subject in this category, it is likely that the data set is too small to draw any significant conclusions about this instrument. Additionally, due to the extremely small and unequal sample sizes when the data is pooled by instrument, it is likely that even non-parametric tests will fail to provide a consistent analysis. In other words, although appearances suggest that the results are not instrument-dependent, it would be necessary to test more subjects to confirm that hypothesis.

4.2 Test 2: Reverberant Energy

4.2.1 Experiment Design

Based on the literature, the 2 variables determined to constitute the parameter of Reverberant Energy were the reverberation time and the direct-to-reverberant energy ratio (D/R). Because ViMiC uses a multi-channel feedback-delay network for its reverberant tail, the user-generated values indicate the percentage of direct sound fed into the feedback loops rather than the length of the reverberant tail. In order to determine the percentages of feedback that corresponded to desired reverberation times, impulse responses were taken of the system and analyzed in Matlab. The impulse responses were taken using a swept-sine signal and 2 Audix TR-40 omnidirectional microphones. Energy decay functions of two sample impulse responses from the system are shown in Figure 14 and Figure 15.

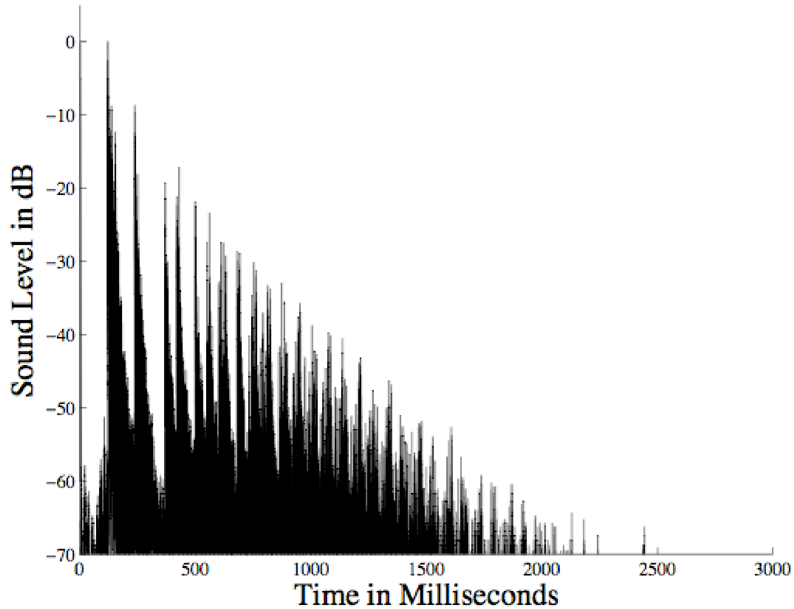


Figure 14: Energy Decay of an Impulse Response from *Jacktrip*/ViMiC Telepresence System. Reverberant Energy Level: -12 dB, % Feedback: 60%.

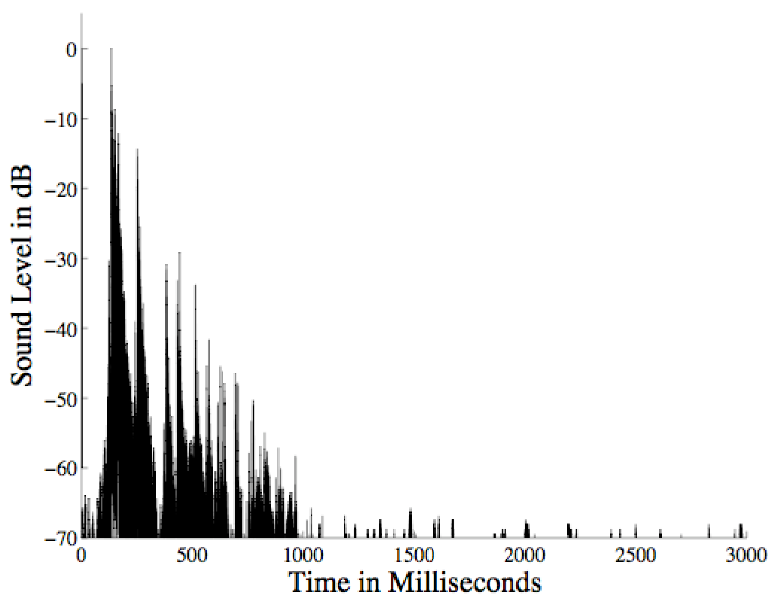


Figure 15: Energy Decay of an Impulse Response from *Jacktrip*/ViMiC Telepresence System. Reverberant Energy Level: -12 dB, % Feedback: 30%.

The results of the Matlab T_{30} calculations are shown in Table 19. The best-fit line is shown in Figure 16.

Table 19: Analysis of ViMiC Reverberation Time (Averaged Over All Frequencies)

0 dB	75%	$T_{30} = 4.09$ s
	30%	$T_{30} = 1.395$ s
	60%	$T_{30} = 2.55$ s
	90%	$T_{30} = 9$ s
12 dB	30%	$T_{30} = 1.795$ s
	60%	$T_{30} = 2.504$ s
	90%	$T_{30} = 7.986$ s
-6 dB	30%	$T_{30} = 0.9546$ s
	60%	$T_{30} = 2.3101$ s
	90%	$T_{30} = 8.9205$ s

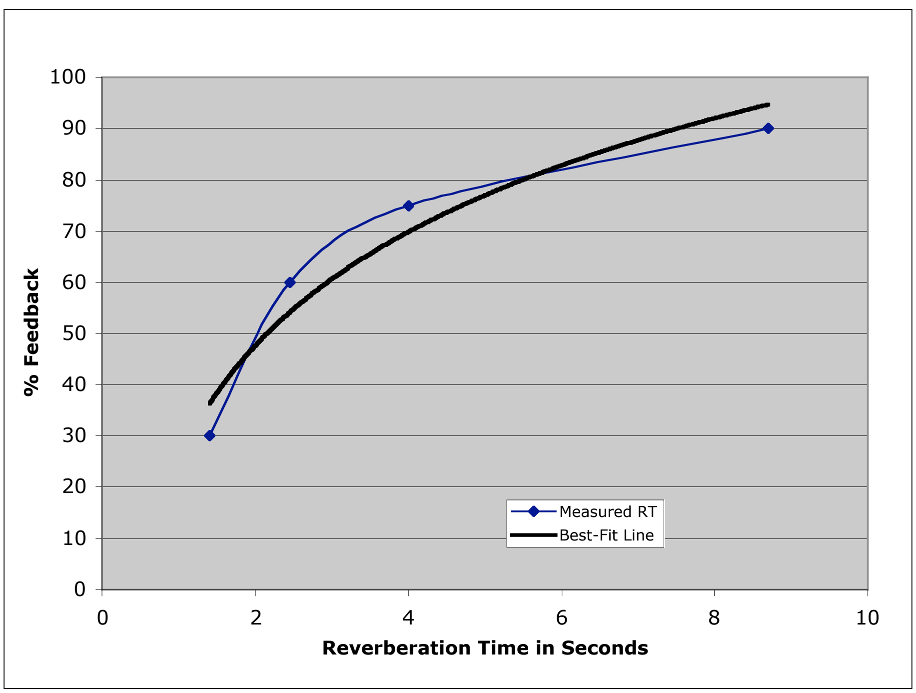


Figure 16: Impulse Response Analysis of ViMiC RT: Best Fit Line

Following these lines, two %-Feedback values were chosen, along with 3 Reverberant energy levels, to create the variables shown in Table 20. All remaining parameters remained constant at the values shown in Table 17. Direct Sound levels remained constant at 0 dB (normal spherical attenuation). In total, each subject participated in 72 trials (12 different musical combinations, 6 different acoustic combinations) so that each acoustic situation was repeated 12 times to reduce experimental error. Complete data sets were collected from all 8 participants.

Table 20: Test #2 Variable Parameters

Reverberation Time (avg. over all frequencies)	0.45 seconds	10% Feedback
	2.1 seconds	50% Feedback
Direct-to-Reverberant Energy Ratio	+12 dB	–12 dB Reverberant Energy Level (REL; below normal spherical attenuation)
	0 dB	0 dB REL (normal spherical attenuation)
	–6 dB	+6 dB REL (above normal spherical attenuation)

4.2.2 Results

The results from all 8 subjects were analyzed by acoustic condition for mean preference and standard deviation, as shown in Figure 17.

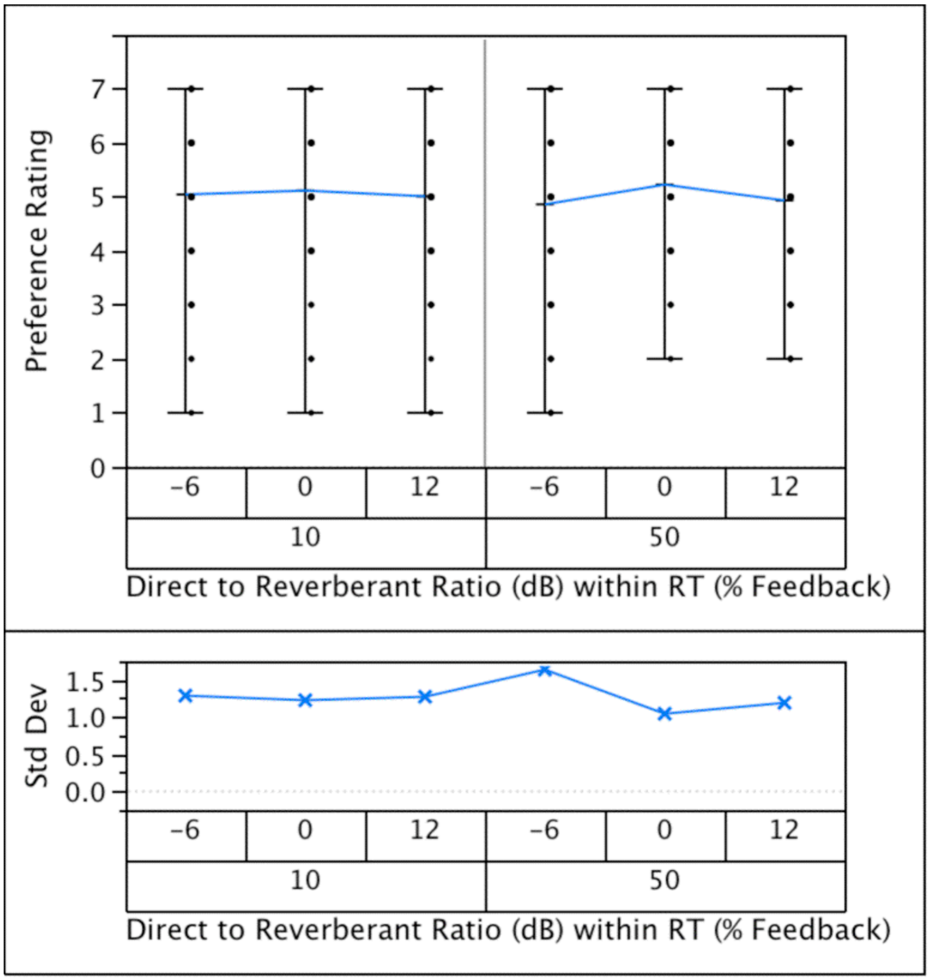


Figure 17: Mean Preferences and Standard Deviation for Reverberant Energy

The abscissa is categorized in two subgroups. The top group shows the Direct-to-Reverberant Ratio (D/R) in dB, and the lower group shows the percentage of feedback corresponding to the two reverberation times listed in Table 20. The ordinate shows user preference rating, with 7 indicating the highest preference/ease of communication and 1 indicating the lowest preference. As the results show, there is no clear trend across all subjects. The slight preference for a D/R of 0 dB and RT of 2.1 s showed no statistical significance in the KS test. In order to determine whether the absence of trends was the same across all subjects or if it was a result of high inter-subject variability, the means and standard deviations were calculated for individual subjects (see Figure 18).

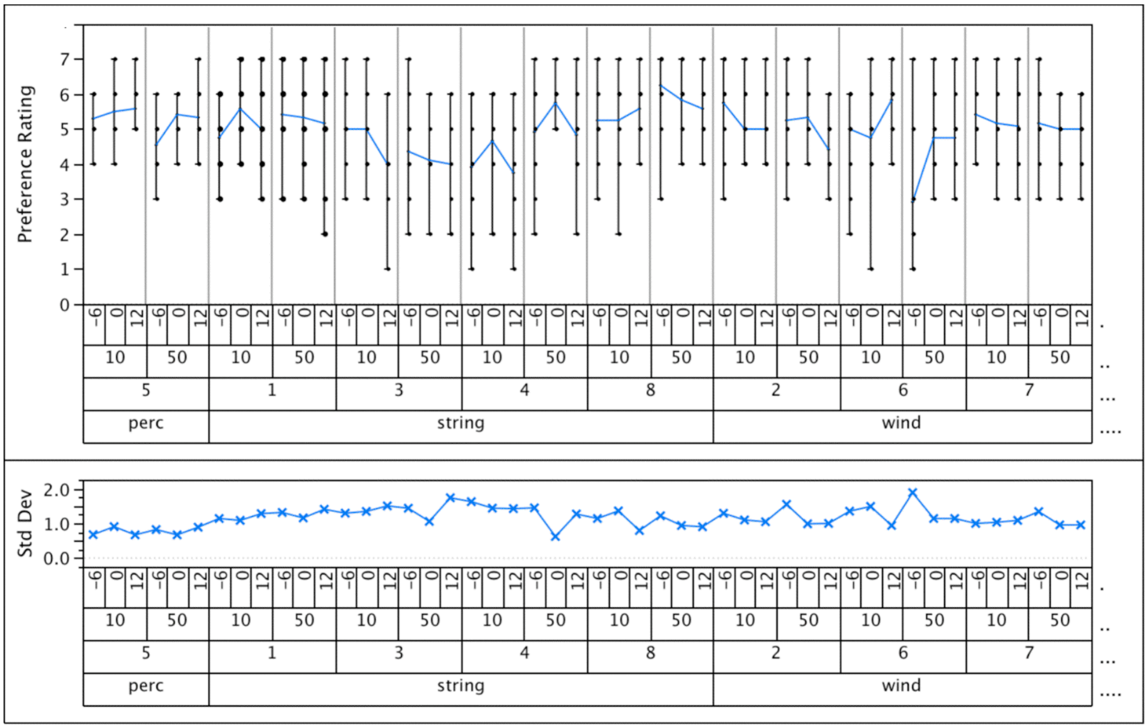


Figure 18: Mean Preference for Reverberant Energy by Individual Subject

As can be seen from the individual preference data, the overall absence of a significant trend can be attributed to wide individual variability. There are 2 distinct groups dividing the Reverberation Time preferences: Group A (Subjects 1, 4 and 8) prefers the longer RT (2.1 s) over the shorter RT (0.45 s). Group B (Subjects 2, 3, 5, and 6) prefers shorter RT over longer RT. Subject 7 had to be eliminated from both groups due to a lack of significant preference for either condition, forming Group C. The mean preferences for RT pooled by group are shown in Figure 19.

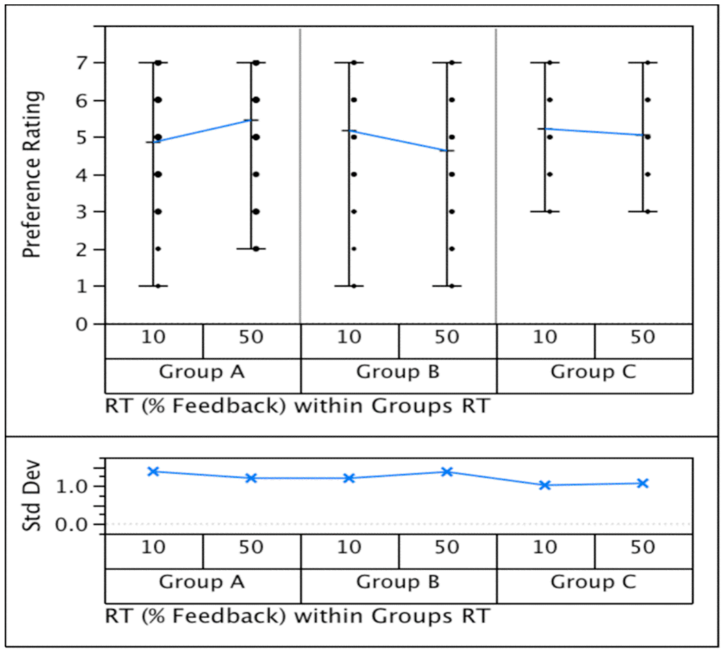


Figure 19: Mean Preference for Reverberation Time Pooled by Group

The KS showed probability *P* values of 0.029 for Group A and 0.0368 for Group B, indicating the significance of these trends. Group C showed no statistical significance. When analyzed by individual subject, the only individual showing statistical significance in RT was Subject 4, with a *P* value of 0.0277. In terms of individual characteristics that might cause such grouping, neither age, experience, nor training correlated with the groups.

There may be some correlation between the groups and the instruments played by the subjects: Group A consisted of all the strings except for Dilruba (which can certainly be considered a different type of instrument from the violins and bass guitar). Group B consisted of the 2 single-reed wind instruments and the 2 instruments with more percussive timbres (Dilruba, percussion). Group C consisted of the only double-reed wind instrument, the oboe, which was also the only female participant. Additionally, Group C followed a similar contour as Group B (both containing wind players), even though the difference was not as significant. One possible explanation is that strings prefer the longer RT because they are used to playing more harmonic roles in an ensemble, long tones which sound better in a reverberant space; on the other hand, instruments with percussive timbres and reeds that are more used to playing virtuosic roles prefer a drier space that does not mask individual shorter notes).

The same groupings did not apply to preferences for Direct-to-Reverberant Ratio. Only one subject showed any statistical significance in preferences for this variable: Subject 6. This subject showed a strong preference for short RT (0.45 s) with a low Reverberant Energy Level (D/R +12 dB), and a strong dislike for long RT (2.1 s) with a high RE Level (D/R –6 dB). This preference could be due to the individual's discomfort with feedback, which occurred more frequently in the most reverberant condition, or simply a preference for less reverberant acoustics. Alone and when grouped with Subject 5 (who followed similar trends), the probability *P* value for any relationship to the most reverberant condition was below 0.0329, and for the relationship between the most and the least reverberant condition (as listed above) the *P* value was 5.29×10^{-4}, or highly significant.

On the other hand, Subjects 1 and 4 show a preference for the 0-dB D/R (normal RE Level) above both strong and weak RE Levels, especially in the condition with longer RT, showing a *P* value of 0.0082. The other 4 subjects (2, 3, 7, and 8) showed no trends in any grouping. What sets Subjects 5 and 6 and Subjects 1 and 4 apart is unclear: neither age, instrument, experience, gender, nor training influenced the groupings. It can only be assumed that these trends are due to individual aesthetic preferences.

For perspective, subjective comments left by the musicians during these tests are shown in Table 21.

Table 21: Subjective Comments: Reverberant Energy Test

Conditions (RT in seconds/ Direct-to-Reverberant Ratio)		Responses
0.45 s	+12 dB	clear, but lack of presence from other performer; anemic (rating: 4)
0.45 s	0 dB	loud enough, but sounded like other performer was in a small closet (rating: 5)
0.45 s	-6 dB	more present (rating: 5)
2.1 s	-6 dB	easy to hear, nice, works well programmatically but not necessarily in all cases (rating: 6/7)
		too present, too much reverb (rating: 2)

A recursive partitioning analysis was done in JMP in an attempt to break down the important factors in this test. In the recursive partitioning done for all subjects, it shows the preference for a middle condition (mid-level RE, 0 dB, is preferred over all; in long RT, low RE Level is preferred, and in short RT, high RE Level is preferred).

Based on the Mean and KS analyses alone, it can be said that different reverberation times are preferred by different groups, most likely by instrument (strings prefer longer reverberation times, winds and percussive instruments prefer shorter reverberation times), which would be useful information when curating a concert of indeterminate music with a specific ensemble in mind. Additionally, in telepresence, it might be possible to create different aural spaces for different instrument groups (although "commonality of aural space" was rated highly in the survey conducted at the beginning of this research). Preferences for D/R appear to be highly individual and unrelated to particular characteristics of the participant. In physical situations, the normal D/R will be easiest to obtain; in telepresence, Reverberant Energy Levels should be kept to normal attenuation due to spherical spreading rather than boosted to add an effect or reduced to add clarity, unless the music specifically calls for such procedures.

4.2.3 Musical Analysis

Because the results of the preference data analysis resulted in few strong conclusions regarding Reverberant Energy, a content analysis was performed on the musical recordings to determine the effects of the acoustic changes. The 1st half of each subject's test (36 trials per person, 8 subjects total) was analyzed based on 26 elements, listed in

Table 22 and organized by category.

Table 22: Elements Used in Musical Content Analysis

Pitch	Volume	Articulation	Ornamentation	Structure	Coordination
Bend	Loud	Rhythmic	Vibrato	Contrast	Tempo
High	Soft	Long Tones	Trill	Counterpoint	Rhythm
Low	Silence	Legato	Tremolo	Call & Response	Style
Glissando	Dynamics	Staccato/Pizz	Volume Pulse	Melody	
		Ext. Tech		Development	

For each trial, the approximate importance/frequency of appearance of each element within the 30-second recording was rated on 0-4 scale. The ratings were 0: none, 1: less than half, 2: half, 3: more than half, and 4: all. Of course, because of the changing music and diverse instructions in the score, it was nearly impossible for a rating of 4 to be given to any element. Ratings of '3' also were quite infrequent for this same reason. The values were then added up and averaged for each acoustic parameter value. The results are shown in Table 24. Because of the rarity of 3 and 4 ratings, it is probably more enlightening to think of the ratings as being on a 0-3 scale. Items in boldface print indicate the highest rated value for each of the two acoustic parameters for the specific musical element.

Table 24: Average Element Rating in Content Analysis: Reverberant Energy Test

	T_{30} **0.45 s**	T_{30} **2.1 s**	**D/R -6 dB**	**D/R 0 dB**	**D/R +12 dB**
Bend	0.74	0.72	0.68	**0.77**	0.74
High	2.07	**2.19**	2.10	2.10	**2.22**
Low	**1.93**	1.81	**1.90**	**1.90**	1.78
Gliss	0.74	0.77	0.79	**0.80**	0.65
Loud	1.85	**2.13**	**2.25**	1.84	1.85
Soft	**2.15**	1.88	1.75	**2.16**	2.15
Silence	**1.21**	0.69	0.83	0.86	**1.23**
Dynamics	1.15	**2.02**	1.50	**1.65**	1.62
Rhythmic	**1.70**	1.17	**1.56**	1.37	1.37
Long Tones	1.25	**1.97**	1.55	**1.65**	1.64
Legato	0.95	**1.97**	1.48	**1.57**	1.29
Staccato./Pizz	**1.32**	0.88	1.05	1.00	**1.29**
Ext. Tech	1.16	**1.37**	1.14	**1.36**	1.31
Vibrato	0.44	**1.65**	**1.07**	1.05	1.03
Trill	**1.20**	0.44	0.81	**0.93**	0.69
Tremolo	**1.22**	0.49	**1.00**	0.81	0.72
Volume Pulse	**1.05**	0.25	0.69	**0.70**	0.53
Contrast	1.26	**1.88**	1.59	1.46	**1.68**
Counterpoint	0.29	**1.97**	1.15	1.04	**1.23**
Call & Response	**1.29**	0.22	0.74	**0.81**	0.72
Melody	0.55	**1.85**	**1.39**	1.05	1.15
Development	0.23	**1.67**	**1.10**	0.88	0.82
Tempo	**1.56**	0.94	1.32	**1.33**	1.05
Rhythm	n/a	n/a	n/a	n/a	n/a
Style	0.18	**1.69**	**1.11**	0.77	0.92

As the average ratings show, there are some visible trends in the performed content across all subjects. An increase in Reverberation Time corresponds to an increase in pitch register, self-produced volume, use of dynamics, long tones, legato playing, extended techniques, vibrato, musical contrast, counterpoint, melodic material, musical development, and stylistic homogeneity. On the other hand, a decrease in Reverberation Time corresponds to a decrease in pitch register, volume, and density, along with the increased use of rhythmic material and staccato/pizzicato playing. Decreased T_{30} also appears to lead to the increased use of tempo coordinations and ornamentations such as tremolos, trills, and volume pulsing, as well as increased call-and-response technique.

Although the changes in content appear to be less consistently affected by changes in Direct-to-Reverberant Ratio, there are a few visible trends. Increased Reverberant Energy Level (decreased D/R) appears to lead to increased stylistic coordination, melodic material and development, vibrato and tremolo as well as increased volume and decreased pitch register. Ornamentations, tempo coordinations and call-and response techniques seem to be used most frequently when the D/R is 0 dB, and with a decrease in REL (increased D/R) comes an decrease in volume and density, increased pitch register and use of staccato/pizzicato playing, as well as an increased use of counterpoint.

Ideally, this content analysis would be performed by additional researchers to eliminate subjective bias and add validity to the data; also, the same tests should be run with additional musical material to eliminate the effects of composer bias. However, this preliminary information may be of use to curators and ensembles interested in creating a concert program that adheres to a specific style. With this information, they may be able to choose a venue with the appropriate acoustics for the playing style they hope to achieve.

4.3 Test 3: Early Reflected Energy

4.3.1 Experiment Design

In this test, two variables were selected. The level of the early reflected energy (after normal attenuation due to spherical spreading was calculated in ViMiC) was varied between 2 levels with relation to the direct sound, which remained constant. Additionally, 3 different room sizes were chosen; however, ViMiC uses 2 different sets of room dimensions for its calculations: 3-variables for rectangular width, length, and height dimensions (in meters) for the 1^{st}-order reflections; and a single-variable room volume (in cubic meters) for the reverberation algorithm. In this test, in spite of the lack of authenticity caused by separation of these two variable sets, the room volume parameter in the reverberation module was held constant, while the X/Y/Z dimension variables were varied based on 3 room sizes. The room sizes were chosen to create the largest possible realistic variation between the initial time delay gaps (ITDG; time in milliseconds between direct sound and 1^{st} wall/ceiling reflection, considered a measure of the subjective impression of "intimacy") in the spaces [2]. The 2 varied parameters and their chosen values are listed in Table 25.

Table 25: Test 2 (Early Reflected Energy) Varied Parameters

	0 dB (normal spherical attenuation)
	-6 dB (below normal spherical attenuation)
Room Dimensions (Width by Depth by Height in meters)	#1: 8 m x 10 m x 5 m (i.e. 400 m^3); ITDG: 26 ms (ceiling)
	#2: 25 m x 20 m x 15 m (i.e. 7,500 m^3); ITDG: 43 ms (ceiling)
	#3: 50 m x 80 m x 30 m (i.e. 120,000 m^3); ITDG: 87 ms (ceiling)

As can be seen from Table 25, 3 vastly different hall sizes have been chosen. Room #1 corresponds to a small bar or art gallery, typical performance venues for contemporary music. In the CATT acoustic model (see Figure 7), plaster absorption data was used for the walls and ceiling, and wood absorption data was used for the floor. This is a common configuration of many art galleries in New York, and led to the choice of cutoff frequencies of 8kHz for the early energy and 6kHz for the late energy used in ViMiC. Additionally, the ITDG of 26 ms is on the cusp of the recommended range used by Marshall in the literature (17-25 ms) [20].

Room #2 corresponds to a small hall or theater, typically used for chamber music. The closest correlate in the literature is Tokyo's Dai-ichi Seimei Hall [2]. Seimei Hall has a volume of 6,800 m^3 (approx. 20m x 26m x 12m), seats 767 people, and has a reverberation time of 1.6 s. Although it is quite rare to see a concert hall or theater with a width-to-depth ratio greater than 1, a few halls in the literature do fit this profile, such as Boettcher Concert Hall in Denver, CO. The choice to have a W/D ratio of 1.25 for Room #2 was an attempt to test the preference for lateral early reflections (in the sagittal plane) vs. those from the transverse plane. Additionally, it is typical to perform in unusual venues and in unusual seating arrangements when larger ensembles are performing contemporary music, and therefore such a high W/D ratio may be less unusual in these situations. The ITDG of 43 ms is outside the recommended range but very close to the preferred reflection delay tested by Gade in his small ensemble tests [20].

Room #3 corresponds to an extremely large space, on the order of a football or basketball stadium. Although such a venue is highly unlikely to be used for indeterminate music performance, there have been documented contemporary performances in unusually large venues (The annual Bang-On-A-Can New Music Festival, as well as performances of Anthony Braxton's piece for 100 Tubas and Stockhausen's *Gruppen* for 3 Orchestras have been held in the World Financial Center Winter Garden Atrium in New York, which has a volume of approximately 184,000 m^3; Roger Reynolds organized a 3-day music festival in 1969 of music by Cage, Gordon Mumma, Toru Takemitsu, and others, that took place in the Yoyogi National Gymnasium in Tokyo)[13]. The closest correlate concert hall in the literature would be Royal Albert Hall in London. Royal Albert Hall has a volume of 86,650 m^3 (approx. 47m x 51m x 36m), seats 5,300 people and has a mid-frequency reverberation time of 2.4 s. Although ceiling reflectors have since improved conditions, the hall was notorious for echoes and lack of support in the past, which is in strong contrast with the other two space types used in this research. Additionally, this size places the ITDG at 87 ms, directly outside the range tested by Gade in the literature as well as just beyond the cutoff time for the C_{80} parameter (80 ms), which would significantly change the value of this parameter, and according to Gade, the impression of "Hearing Each Other" in the space [20].

The remaining parameters were held constant at the values shown in Table 17. Direct Sound Level was fixed at 0 dB (normal spherical attenuation).

---

[13] http://www.xs4all.nl/~cagecomp/1912-1971.htm

4.3.2 Results

Due to technical difficulties only six trials were recorded for one of the subjects, resulting in 7 complete data sets of 72 trials each for this experiment. The mean and standard deviation for each condition across all subjects are shown in Figure 20. The abscissa separates the mean preference for each Room Dimension (1-3) within the two values for Early Reflected Energy Level. The ordinate represents user preference for ease of communication in each condition, 7 being the highest preference and 1 being the lowest preference.

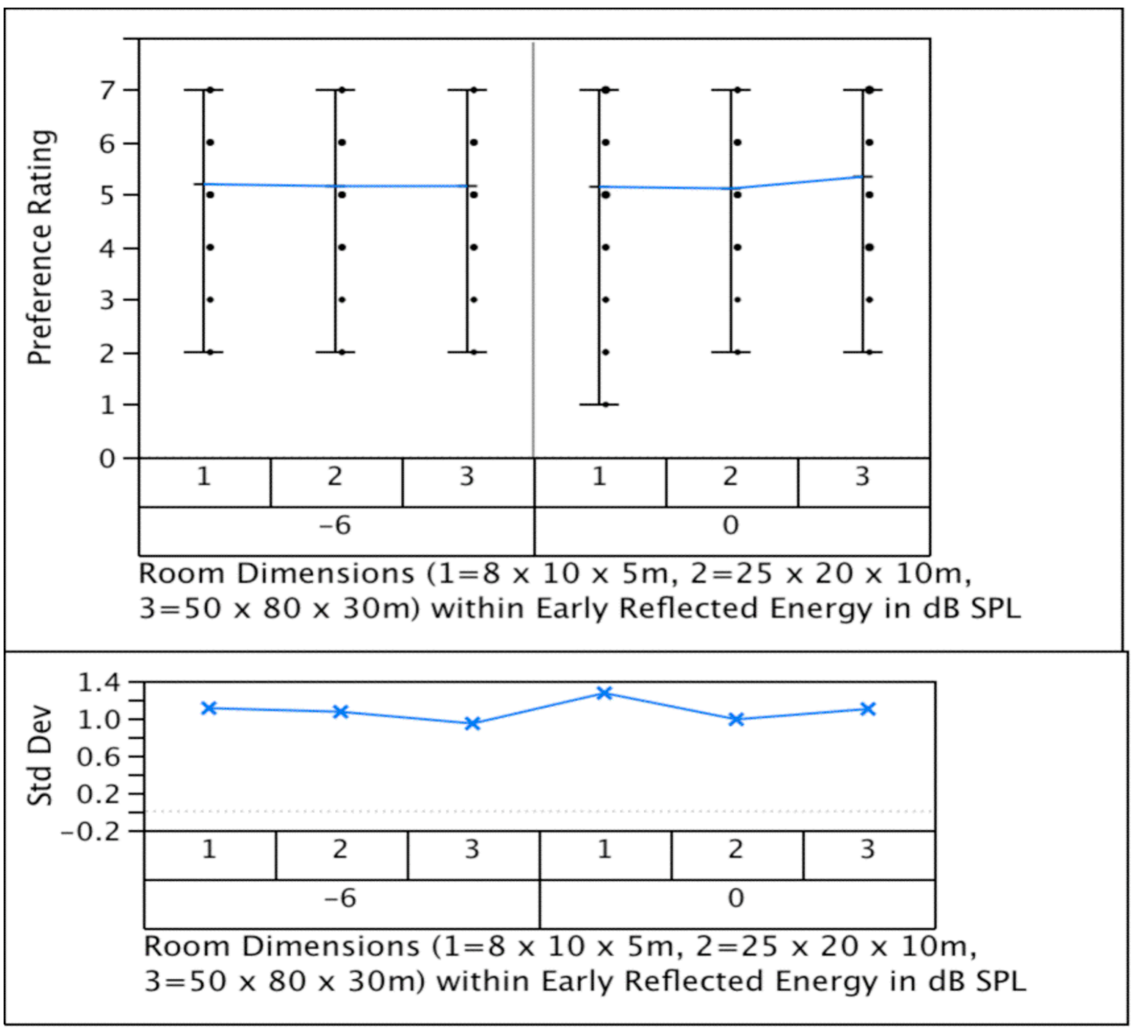


Figure 20: Mean Preference Ratings for Early Reflected Energy

For the first Early Reflected Energy Level value (-6 dB), user preference appears to be high on average, with a constant mean slightly above 5. However, at this level there appears to be no change in preference from one room dimension to the next. At a level of 0 dB, the preference is also high but with very little change, although the largest room (#3) is rated slightly higher on average and the mid-sized room slightly lower (#1 = 5.2, #2 = 5.1, #3 = 5.4).

The KS test was performed on all condition pairs across all subjects to determine the significance of these apparent trends. It was found that no combination across all subjects was rendered statistically significant (in fact, probability P never fell below 90%). The subjects were pooled based on similar within-subject distributions and the results were still not found to be statistically significant. To determine the cause of this, the means and standard deviations were calculated for each individual subject, as shown in Figure 21.

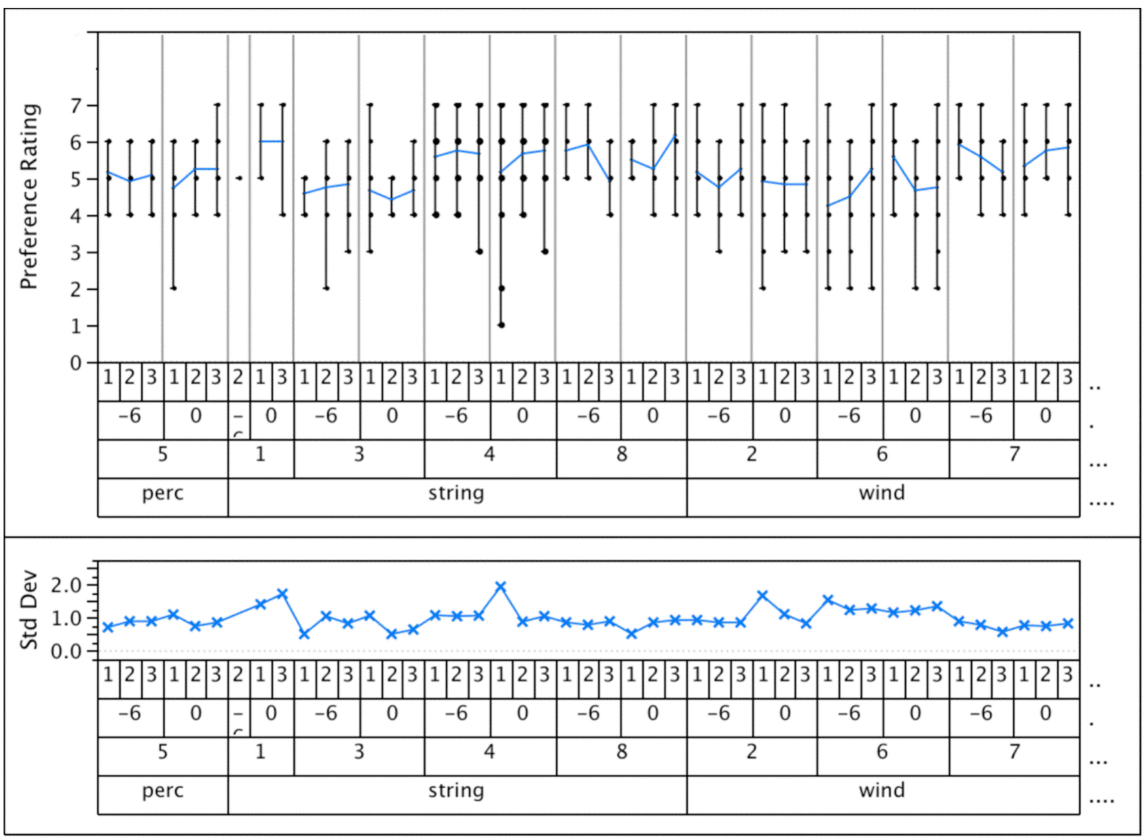


Figure 21: Mean Preferences for Early Reflected Energy (ERE) by Subject

As can be seen from Figure 21, the variation is great between subjects and explains the lack of significant trends in the group mean comparisons. There are no clear instrument-dependent trends, either. Subjects 3-5 do not show any trends when ERE Levels are –6 dB. At this level, Subjects 7 and 8 prefer Room #3 the least, Subject 2 prefers Room #2 the least, and Subject 6 prefers Room #1 the least. Trends are more visible when ERE Levels are 0 dB. In this situation, Subjects 4, 5 and 7 followed the same trend, preferring Room #3 the most and Room #1 the least. Subjects 3, 6 and 8 preferred Room #2 the least (Subject 6 prefers Room #1 the most, Subject 8 prefers Room #3 the most), and subject 2 shows no specific preference. Subjects 6-8 showed complete opposite trends for the 2 different ERE levels. No known individual characteristics explain these groupings (individuals are not pooled based on instrument, experience, age, or training). Additionally, the KS test did not find significance in any of the aforementioned trends. Therefore, it must be concluded that either not enough subjects have been tested to find any trends significant, technical inaccuracies or the effect of low Reverberant Energy (~0.45 s) are skewing the data, or that Early Reflected Energy is not an important variable in musician communication for indeterminate music. Due to the high individual variability and the lack of significant trends for any individual subject, it seems likely that the latter is the case.

It seems that room size (and the corresponding ITDG) becomes more important when the reflections are more audible. The lack of linear progression from Room 1-3 in most subjects shows that the W/D ratio (<1 in Rooms 1 and 3, >1 in Room 2) may be important, whether its effect is positive or negative. Additionally, Room 2 has an ITDG of 43 ms, which in Gade's research was found to be ideal. Therefore, it may be the most familiar to these subjects, and their impressions of the condition may be affected by their experience in such a space. On the other hand, it is interesting that some subjects preferred Room 3, with an 87 ms ITDG, because this would presumably mask fast musical passages. It is possible that due to the low Reverberant Energy, the Early Reflected Energy in this range is adding to an impression of reverberance, which is preferred by some musicians. This may indicate (in combination with the results of Test 2) that Reverberant Energy is more important than Early Reflected Energy in indeterminate music, a conclusion that is opposite to that found by Gade for classical music. It is also interesting that some subjects rated Room 1 as least preferred, considering that this room is the most common venue size for indeterminate music performance. However, all conditions in this test were rated highly on average (means scores all above 5) so all that can be said conclusively about this test is that changes in Early Reflected Energy are either inaudible or subject to large individual variation, a result that is very different from results found when testing classical music. The results of each test are compared with each other to determine which parameter was most important in the conclusions section of this paper.

4.3.3 Musical Analysis

Due to the lack of significant conclusions drawn from the Early Reflected Energy experiments, a musical content analysis was performed on the audio recordings from the 1st 36 trials of each subject (see Section 4.2.3 for full explanation). The musical elements chosen for analysis are listed in

Table 22. The results of the analysis for Early Reflected Energy are shown in Table 26. Again, it is more useful to think of the ratings as 0-2, and the items in boldface print are the parameter values most correlated to that musical element.

Table 26: Average Element Rating in Music Content Analysis

	REL 0 dB	REL –6 dB	Room #1	Room #2	Room #3
Bend	**0.78**	0.67	**0.83**	0.71	0.61
High	**2.18**	1.96	2.11	**2.16**	1.92
Low	1.67	**1.92**	1.83	1.70	**1.93**
Gliss	**0.68**	0.62	0.65	**0.71**	0.59
Loud	**2.18**	2.03	2.25	**2.29**	1.71
Soft	1.68	**1.77**	1.64	1.58	**2.04**
Silence	0.88	**1.01**	0.71	0.80	**1.42**
Dynamics	**1.69**	1.47	1.67	**1.71**	1.30
Rhythmic	1.53	1.53	**1.74**	1.63	1.19
Long Tones	**1.66**	1.55	1.45	1.62	**1.79**
Legato	1.25	**1.35**	0.97	1.36	**1.69**
Staccato./Pizz	1.34	**1.43**	1.47	**1.49**	1.20
Ext. Tech	**1.65**	1.31	**1.70**	1.49	1.15
Vibrato	0.52	**0.72**	0.50	0.64	**0.79**
Trill	**1.21**	0.98	**1.29**	1.07	0.85
Tremolo	**1.12**	1.00	1.06	**1.11**	1.00
Volume Pulse	**1.35**	1.11	**1.49**	1.22	0.90
Contrast	**1.37**	1.27	**1.47**	**1.47**	0.91
Counterpoint	0.44	**0.72**	0.65	0.64	**0.74**
Call & Response	**1.14**	1.04	**1.47**	1.13	0.92
Melody	0.65	**0.98**	0.65	0.90	**1.09**
Development	0.50	**0.70**	**1.47**	0.70	0.79
Tempo	**1.41**	1.27	0.65	**1.26**	1.12
Rhythm	0.97	**1.21**	**1.47**	1.37	1.14
Style	0.46	**0.64**	0.65	0.50	**0.99**

As shown in the ratings, an increase in room size (corresponding to an increase in ITDG) corresponds to a lowering in pitch register, decrease in self-produced volume, the presence of silence, long tones, legato playing, vibrato, counterpoint, melodic material, and tempo and stylistic coordination. On the other hand, the decrease in room size may lead to an increase in ornamentations such as pitch bends, glissandi, extended techniques, trills, and volume pulsing as well as increased volume and density and use of higher pitch register. The decreased room size also corresponds to an increased use of rhythmic material for coordination, use of a broader range of dynamics and contrast, and call and response patterns between the musicians. Interestingly, although the increased room size appears to lead to increased melodic material, the decreased room size corresponds to increased musical development (although the interpretation of what constitutes "development" is highly subjective). This implies that neither space has particularly negative connotations for performance of indeterminate music, just that an ensemble wishing to create a specific style of music could use this research as a guideline for choosing the correct venue.

Additionally, an increase in Early Reflected Energy Level appears to lead to increased pitch register, self-produced volume, use of dynamics, musical contrast, trills, tremolos, volume pulsing, call and response, tempo coordinations; as well as a lack of melody, development and counterpoint. The opposite is true of a decrease in level. The remaining elements appear to be more influenced by room-size changes than by level changes.

As a caveat, however, it should be noted that these results may be specific to the Wolff piece alone and other material would need to be tested before drawing any strong conclusions from the two musical analyses.

4.4 Test #4: Sound Source Position/Visual Connection

4.4.1 Experiment Design

The final test contained values for 2 parameters: the position of the virtual sound source within the 2-dimensional horizontal plane. X and Y coordinates were calculated based on the user-generated angle (between sagittal plane of musician and virtual source position).

The equations are listed below:

$$X = x \bullet \sin\left(\frac{\pi\theta}{180}\right)$$

$$Y = x \bullet \cos\left(\frac{\pi\theta}{180}\right)$$

where x is the virtual source distance in meters [6].

The values are listed in Table 27. For the sound source position, omnidirectional source directivity was assumed.

Table 27: Variable Parameters for Test #4

Sound Source Position	0º (front)
	90º (left side)
	180º (rear)
Visual Connection	0 (video off)
	1 (video on)

For the visual connection, a black pop-up window was generated in Pure Data that would cover the video image of the remote musician with a black screen. This procedure was chosen because any command that would turn off the video cameras or sever the digital video connection would add significant latency to the process, whereas this procedure promised to be quick and reduce processing requirements. However, some technical difficulties with this command on the Linux Terminal in Room B caused one subject's data to be unusable for this test. After this experience, the remaining subjects were given all trials with video on in the first half of the test and trials with video off in the second half of the test. The remaining parameters were held constant using the values listed in Table 17. Direct Sound was transmitted at a level of 0 dB (normal spherical attenuation). In total, 7 complete sets of data were collected, with 72 trials in each set.

4.4.2 Results

The results from Test #4 were analyzed by mean preference and standard deviation across all subjects, and the analysis is shown in Figure 22.

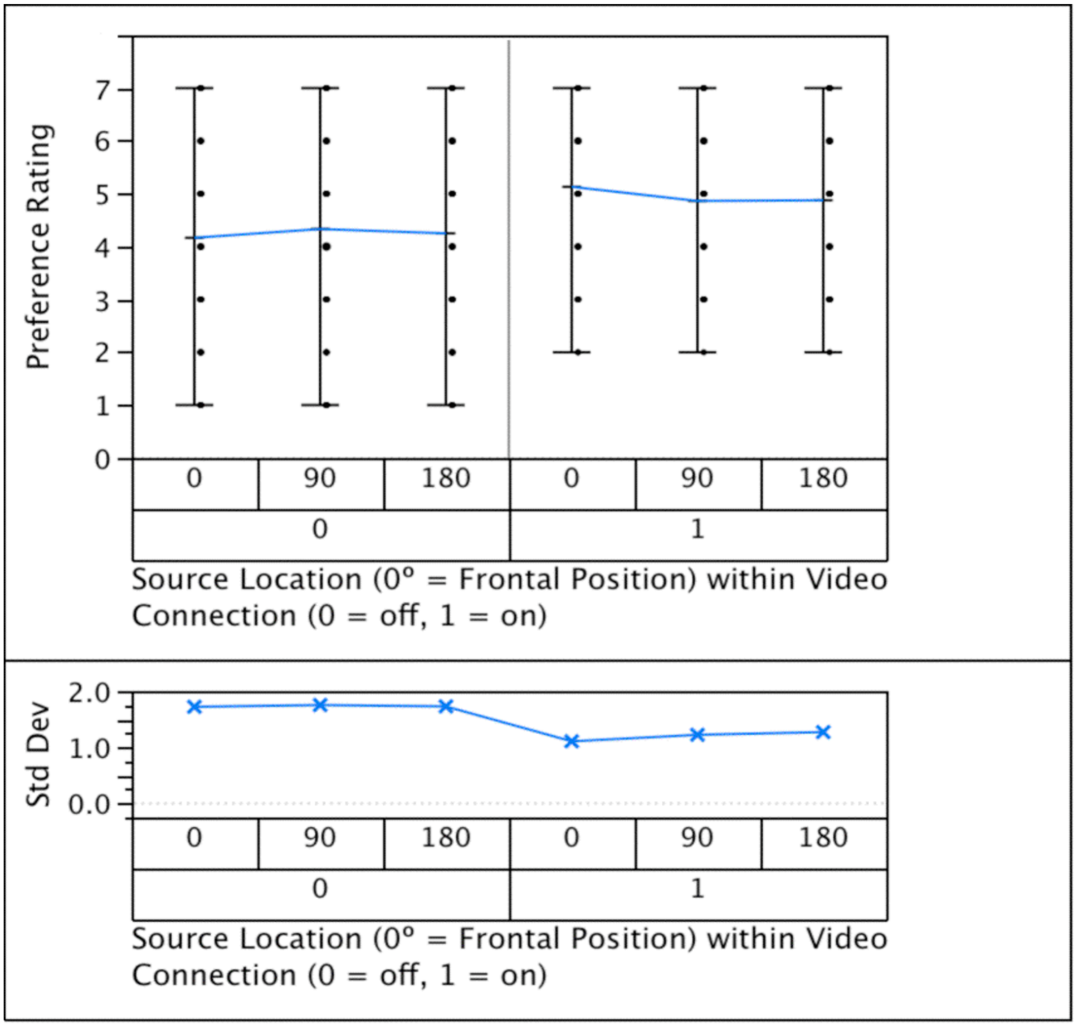


Figure 22: Mean Preference for Sound Source position and Visual Connection

The abscissa of Figure 22 has two subsets. The top subset lists the angle of the remote virtually-located musician in degrees with respect to the present musician's sagittal plane. The lower subset lists the presence of visual image of the remote musician, 1 meaning full-body visual image (no blocking window) and 0 meaning a lack of any visual image (the presence of a blocking window). The ordinate shows user preference (7 indicating highest preference/ease of communication, 1 indicating lowest preference).

From this graph, it appears that the presence/absence of video has a significant effect on preference (presence of visuals is preferred). Figure 23 shows the mean preference and standard deviation for visual connection across all subjects across all sound source position trials.

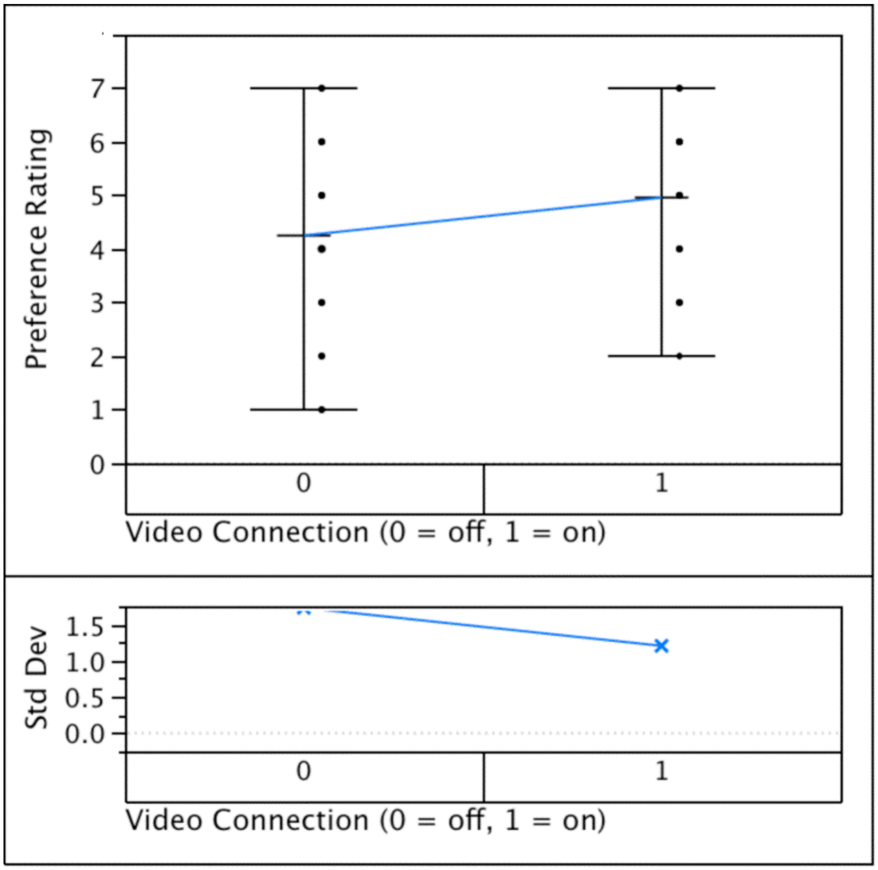


Figure 23: Mean Preference for Visual Connection

A significant trend is apparent in the graph: although the individual variability appears to increase in the absence of the visual image, the average ease of communication appears to decrease. A KS test confirmed the statistical significance of this trend, with a probability *P* value of 3.1 x 10^{-5} (<<0.05).

From Figure 22 it appears that the change in source position has a very limited effect on preference, although there appears to be a slight preference for 0º in the presence of video and an even slighter preference for 90º in the absence of video. The mean preference and standard deviation for source position across all visual conditions and all subjects is shown in Figure 24.

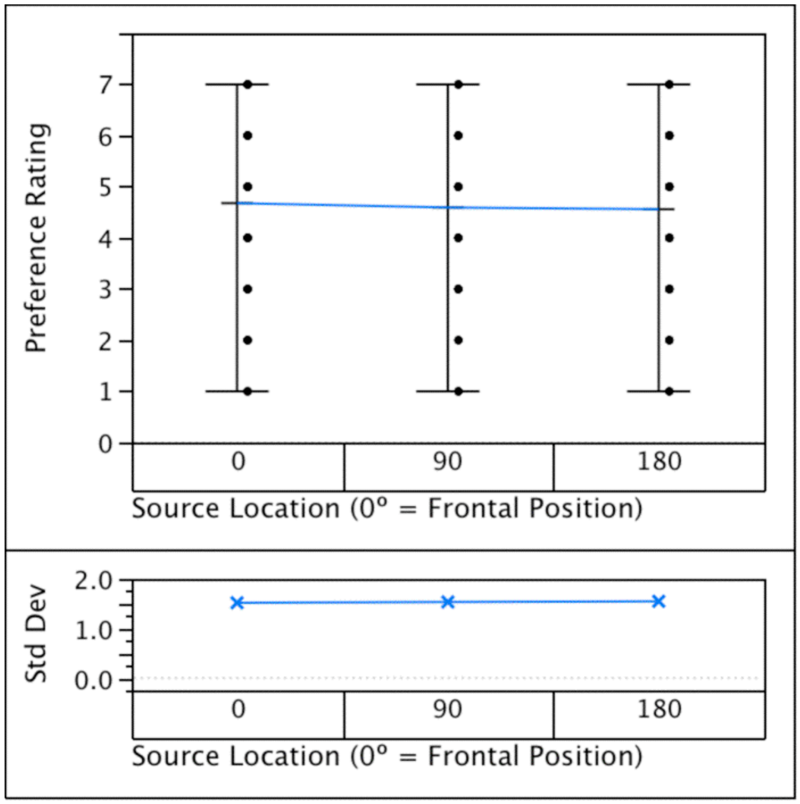


Figure 24: Mean Preference for Sound Source Position

Again it can be seen that changes in source position have a very limited effect on preference, although 0º is most preferred and 180º least preferred. The KS test shows no statistical significance of this data, when grouped across all trials or separated by presence of visual image, or when pooled by instrument group. To determine whether this trend was common across all subjects, the mean preferences and standard deviations were calculated for each individual subject (see Figure 25).

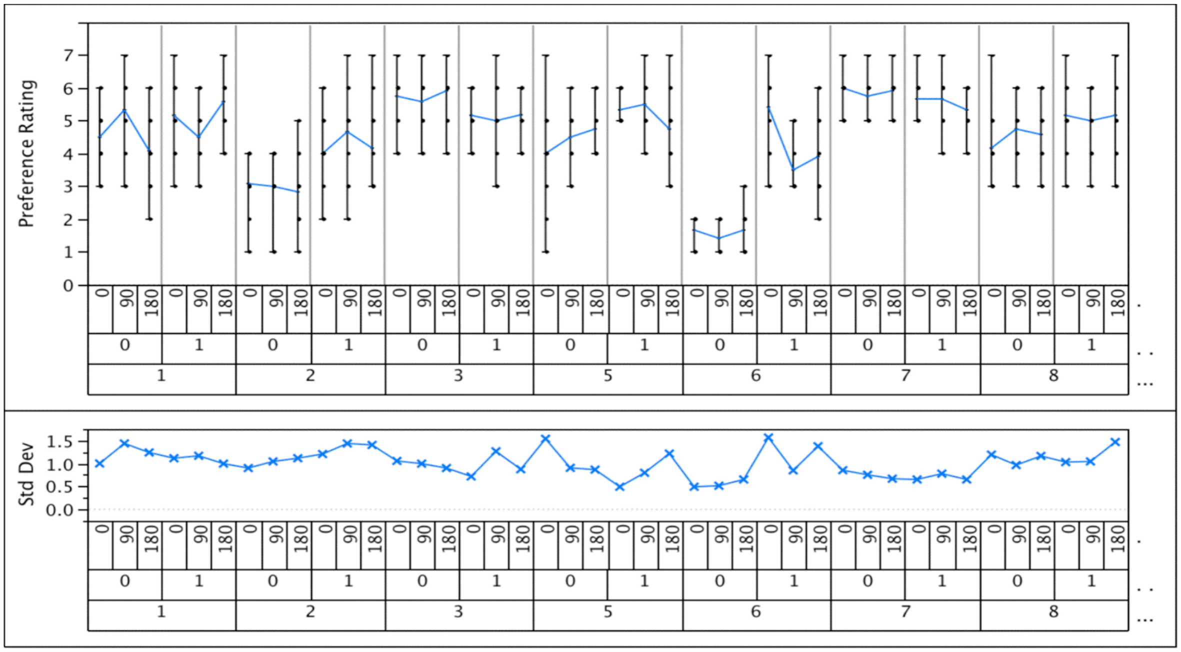


Figure 25: Mean Preference for Test #4 by Individual Subject

The wide variability between subjects is clearly visible. Subjects 1 and 8 are similar in preferring front and rear positions in the presence of video and preferring the side position in the absence of video (interestingly, these two subjects were partners for this test). Subjects 3 and 6 (also test partners) were similar in that they least preferred the side position of the virtual sound source regardless of the presence or absence of visual. In fact, Subject 6 so adamantly disliked the side position that it was confirmed through subjective comments after the test and the change in preference between this Subject's ratings for 0º and 90º in the presence of visual image was the only source-position within-subject relationship found to have statistical significance ($P<0.05$). Subjects 2 and 5 preferred 90º in the presence of video, Subject 2 preferred 0º in the absence of video and Subject 5 preferred 180º in the absence of video. Subject 7 preferred both 0º and 90º in the presence of video and preferred 90º the least in the absence of video.

There are a few possible explanations for these choices:

1. In the presence of the visual image, it seems logical that the user would prefer a sound source position that provides greatest synchrony with the visual image (0º). This explains why Subjects 1, 3, 6, and 8 prefer 90º the least (it is possible that 180º, because of the similar Inter-aural Level and Inter-aural Time differences to those at 0º, seems less

jarring in the presence of a 0º visual image than 90º). It could also explain why subjects 5 and 7 prefer 180º the least (it is the farthest away from 0º, and in the presence of a ventriloquist effect, 90º may be integrated close to 0º, depending on the individual). To explain why Subjects 2 and 5 prefer the 90º-position in the presence of visuals, it may be necessary to use a different explanation:

2. Depending on the instrument directivity and range, the position that would provide the greatest contrast with the self-produced sound (and therefore the least masking) would be preferred [41]. Starting with the first pair of subjects, Subject 1 played bass guitar with a large Peavey amplifier on the floor forward and to the left of the musician (facing forward). Subject 8 played violin (instrument positioned to the upper left of the musician). With the majority of bass guitar fundamentals below 100 Hz, the directivity of the amplifier at such frequencies is likely to be fairly uniform. However, it is expected that some lobing will occur at higher transient and formant frequencies.

Violin fundamentals begin around 250 Hz, with formants around 400 Hz, 1 kHz and 3 kHz [41]. Violin directivity at low frequencies is almost omnidirectional; as it increases in frequency the directivity narrows with increasing intensity directed towards the right side of the musician. Additionally, it is shown in the literature that, around frequencies of 1 kHz, sound from the lateral directions is least masked by violin sound [42] (may indicate that the transient frequencies of the bass guitar are more important for ease of communication than low fundamental frequencies, and explain the preference for 90º position). This may also explain the bass player's preference for a 90º position: the expected directivity for the violinist at the least-masked higher frequencies is to the right of the musician, or to the left of the bass player. This would explain the preference for the 90º position. Unfortunately, the data for the other participating violinist was unusable so a comparison cannot be made.

Subject 3 also played bass guitar in the same configuration. However, Subject 5 played soprano saxophone, which is assumed to have a similar directivity as clarinet. The directivity of this instrument is concentrated toward the front of the player at almost all frequencies, which would explain why Subject 3 least prefers the side position, since it is expected that a reed instrument will radiate from 0º.

As for why Subject 5 (saxophone) preferred 0º least (as well as Subject 7, who played oboe), and why Subject 2 (clarinet) preferred 0º most, it may support a hypothesis (supported by the fact that instruments with side-radiating instruments preferred side positions) that in this type of music, a position that allows a blending of the two instruments is preferred. Therefore, the location of greatest self-produced energy is the location most desired for the greatest intensity of transmitted sound (when the visual image is not asynchronous). This is interesting because it implies that masking is actually desired in this type of music because a blending of timbre into a single sonority is preferred.

Technical difficulties and the lack of statistical significance of any of the abovementioned data cast doubt on such a hypothesis. The first issue is that feedback between the microphone and speakers caused some self-produced sound to reproduce in the same room at the speakers along with the remote sound. Therefore, for a musician playing an instrument with strong 0º intensity, a transmitted sound position of 0º may be preferred simply because some of that musician's self-produced sound will also be coming out of the speakers at 0º.

Additionally, the sound reproduction method is limited to 2 dimensions and 45º increments, leading to highly simplified and perceptually inaccurate spatialization. Therefore, it is possible that with more accurate and precise spatialization techniques (Ambisonic or Wave-Field Synthesis reproduction), these judgments would show more significant trends. In future work, more accurate spatialization techniques will be combined with synthetic source directivities to obtain more accurate judgments on this parameter.

However, it can be said that simply placing musicians side-by-side or face-to-face as is often the case in indeterminate music (carry-over from traditional chamber and symphonic ensemble arrangements), a more complex set of guidelines should be followed, relating to the specific directivities and frequency ranges of the participating instruments, and the level of masking desired. Additionally, the presence of visual image or connection with the remote musician is a necessity for ideal musical communication in indeterminate music.

5. Conclusions

5.1 Global Preference Ratings: Comparison

The mean highest-rated and lowest-rated values for each acoustic parameter tested in this research were compiled into two graphs for comparison. The results are shown in Figure 26.

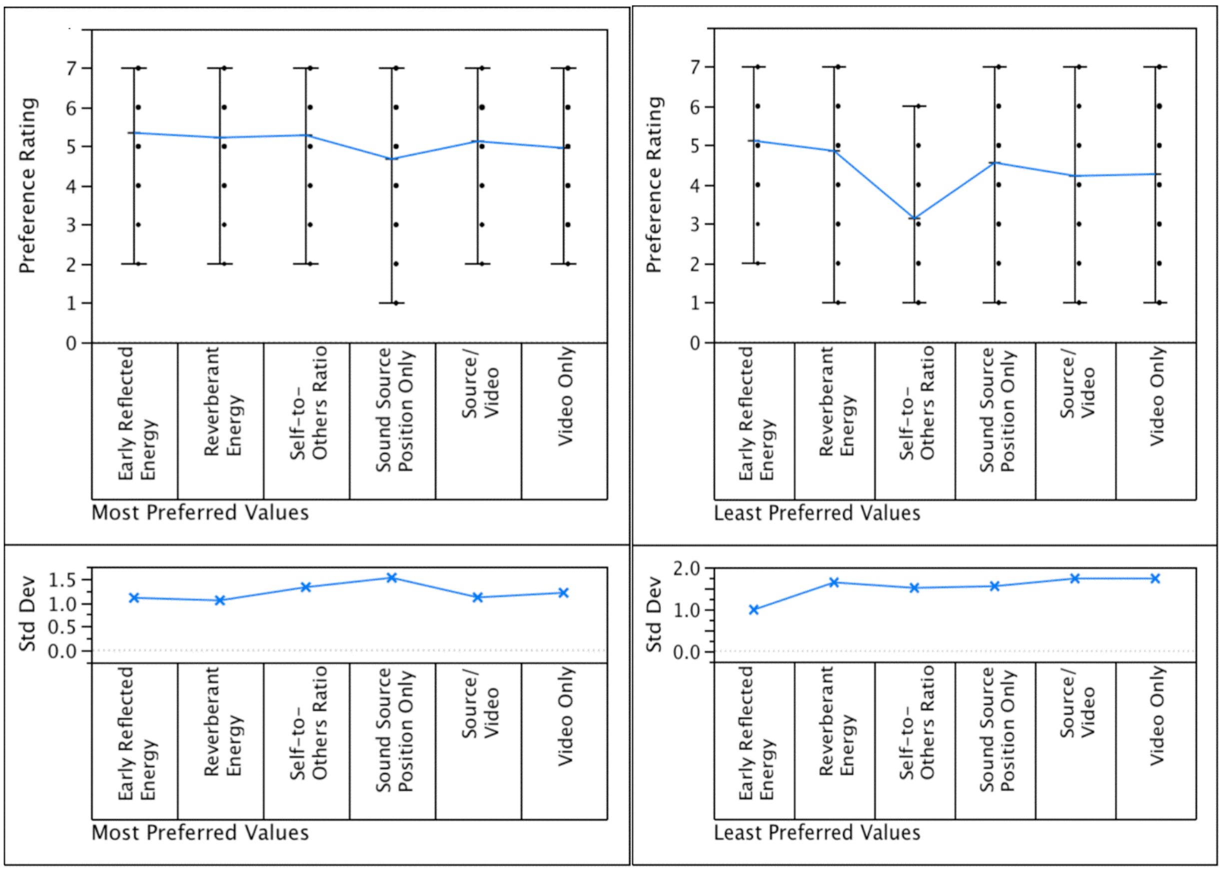


Figure 26: Highest-Rated and Lowest-Rated Values for All Tested Parameters

The difference between the mean preferences for the highest-rated values of all parameters is slight, and it would be difficult to draw any significant conclusions from this comparison. However, when looking at the mean preferences for the lowest-rated parameter values, some trends are visible. According to this comparison, degradation of Self-to-Others Ratio (+18 dB) has the strongest negative effect on communication in indeterminate music, followed by the loss of Visual Connection, then a shift (180º) in Sound Source Position, an excess of Reverberant Energy (T_{30} 2.1s, D/R –6 dB), and lastly by an excess of late Early Reflected Energy (0 dB, Room Size: 50 m x 80 m x 30 m).

A KS test was performed on all parameter relationships for the lowest-rated value data. All relationships were found to have statistical significance except the following three differences: Reverberant Energy/Sound Source Position; Reverberant Energy/Visual Connection; Sound Source Position/Visual Connection. Therefore, it can be said that while no single parameter significantly improves communication, the degradation of certain individual parameters can have a strong negative influence on communication quality.

From these comparisons, it can safely be said that the Self-to-Others Ratio is the most important parameter when complex communication is crucial to the performance. A transmitted level of up to 12 dB more than the self-produced volume is desired, and a levels below 0 dB cause above-average communication difficulties. The Visual Connection is the 2nd most important parameter (full visual clarity is necessary for ease of communication). Sound Source Position (0º preferred) and Reverberant Energy (T_{30} 2.1 s, D/R 0 dB preferred) are similar in importance, but both are considered significantly more important than Early Reflected Energy. Individual variability in preferences for the last 3 parameters indicates that these preferences may be more aesthetic than functional.

5.2 Discussion: Psychoacoustics

While the applications of this research may be self-explanatory in the fields of stage acoustics and telepresence research, the connection between this research and psychoacoustics may seem tenuous. This section will attempt to draw a line between the results of this research and the psychoacoustics literature described in Section 2.2 of this paper.

Real-world situations require more complex auditory scene analysis and auditory stream segregation than laboratory situations [44]. This research may be helpful in understanding how such processes are affected and how musicians adapt in complex situations – for example, individual variability in some parameters may indicate that different musicians prioritize different parameters in auditory stream segregation

[14,31]. In terms of higher-order segregation (*Gestalt*s), the observed trends in musical content may indicate a process of active assimilation (changing the content to fit a pre-existing schema) based on previous experience with the acoustics of specific venues [30,66].

The hierarchical relationships between parameters may be viewed through the lens of Jekosch's description of semiotic cues necessary for assigning meaning to the communicated musical material [30]. In the presence of specific acoustic conditions that are both preferred and draw out stylistic homogeneity/counterpoint, this may indicate a situation where semiosis is prolific between musicians. However, as the preference ratings do not always reflect the changes in content, this may indicate that in indeterminate music, auditory scene analysis and semiosis are less well-defined. Musical objects may be kept permanently in the undifferentiated foreground and the ability to assign clear meaning may not predicate successful communication in indeterminate performance [66]. In terms of cross-modal effects, the shifting of preferred sound source position in the presence of visual connection shows that preferences can be affected by multi-modal information [37,54].

To return lastly to Möller's concept of Human-Human Interactions [43], it can be determined based on the statistical significance of certain parameters, Self-to-Others Ratio and Visual Connection have the most direct connection to complex communication quality because there is such a clear consensus for all users on quality and acceptability. As for the other 3 parameters, the high level of individual variability indicates that their importance to complex communication may be less functional and more aesthetic, as mentioned above.

5.3 Future Work

As outlined in previous sections, many possible avenues of future research may be necessary to augment the results shown here. Additional subjects would add validity to this data, possibly covering a wider range of instrumentation. Sound-reproduction setups including elevation and real-time convolution with real or CATT-Acoustic impulse responses would also add to the spatial accuracy of the experience, and eliminate some possible sources of experimental error. A control experiment, connecting the 2 locations with a cat5 cable instead of internet transmission, would be useful to pinpoint the sources of error in the internet transmission itself. A 2AFC test designed for multi-dimensional scaling might be more useful, as well, in light of the high individual variability allowed by the preference ratings. Different musical material (and possibly unguided improvisation sessions) should be added to eliminate bias.

Along with the system improvements described above, more parameters should be tested, such as source directivity, amount and location of diffusive material, masking (additional musicians in each room), background noise, time delay, frequency content of reflections and reverberant tail, and the presence of visual images of the space. In particular, the masking experiments would further illuminate the connections between this research and Auditory Scene Analysis. More complex acoustic situations such as directivity and diffusion especially would benefit from impulse responses generated in CATT-Acoustic.

Ideally, other forms of complex communication may be tested along with indeterminate music. Spoken conversation and musical education would both be useful scenarios to test using this setup. Finally, a Quality Assessment Index (similar to Möller's concept) could be developed to predict, based on the geometrical/material data from the space, the communication quality and best scenario (type of communication, arrangement of musicians or speakers) for communication in that space.

5.4 Summary

In this work, research has been conducted into the relevance of 5 acoustic parameters to musician communication in indeterminate music, a form of complex communication. In indeterminate music, elements of the score are left open to the performer, creating a situation that is highly sensitive to the acoustics of the performance space [46]. A literature review of classical stage acoustics, contemporary music theory, telepresence research, communication acoustics and psychoacoustics provided the foundations for this research. Previous experiments by Gade and Ueno [20,63] provided the structure for these experiments, along with discussions obtained from pilot experiments and a survey completed by 9 contemporary musicians. In these experiments, 2 musicians were visually and acoustically isolated, connected via internet transmission, in a situation known as telepresence [6]. The musicians attempted to perform excerpts from Christian Wolff's *For 1, 2, or 3 People* (1964) and rate their ease of communication in each acoustic condition. Within this arrangement, 4 tests were conducted, varying different parameters for each test.

From individual preference ratings and musical content analysis, it was determined that the most important parameter for complex communication was the Self-to-Others Ratio. A high level of transmitted sound was desired (a condition difficult to create in physical reality but simple and flexible in the realm of telepresence), resulting in Self-to-Others Ratios as low as –12 dB. The second most important parameter was the Visual Connection between the musicians. The absence of video was determined to have a significantly detrimental effect on the musicians' ability to communicate in this type of music. Reverberant Energy and Sound Source Position had similar significance, although Reverberant Energy was subject to high individual variability, and was found to have significance when the participants were pooled into groups, roughly dictated by their instrument. One group (winds and percussive instruments) preferred less reverberant spaces, whereas the other group (strings) preferred more reverberant spaces. Sound Source Position was subject to a lower level of individual variability. Some individuals found the side position to be highly distracting and preferred the front position, while others preferred the side position to front and rear positions. One possible explanation for this difference may be the instrument-dependent directivity characteristics of each instrument, and the subsequent frequency-dependent masking of musician pairs.

Early Reflected Energy was determined to have the least importance for communication efficiency. However, a musical content analysis of musician performances in conditions with varying Early Reflected Energy and Reverberant Energy showed that although individual variability and lack of significant trends may have proved inconclusive for the preference ratings, some trends were visible in the performed content.

An increase in Reverberation Time was found to correspond to an increase in pitch register, self-produced volume, musical contrast, counterpoint, melodic development, and stylistic homogeneity, among other things. On the other hand, a decrease in

Reverberation Time corresponds to a decrease in density, increased use of rhythmic material and call-and-response techniques. A decrease in Direct-to-Reverberant Energy Ratio appears to lead to increased stylistic coordination, melodic development, vibrato, and increased volume and decreased pitch register. An increase in D/R corresponds to a decrease in volume and density, increased pitch register and use of staccato/pizzicato playing, as well as an increased use of counterpoint.

An increase in room size with regards to the Early Reflected Energy (indicating an increase in ITDG [2]) corresponds to a decrease in self-produced volume and density, increased legato playing, vibrato, counterpoint, melodic material, and tempo and stylistic coordination. On the other hand, the decrease in room size may lead to an increase in use of ornamentations, rhythmic material for coordination, a broader range of dynamics and contrast, and call and response patterns between the musicians. Additionally, an increase in Early Reflected Energy Level appears to lead to increased pitch register, self-produced volume, use of dynamics, musical contrast, ornamentations, call and response, and tempo coordination, as well as a lack of melody, development and counterpoint; the opposite is true of a decrease in ERE Level. This information could be used by ensembles choosing concert venues based on the style of improvised or indeterminate music they hope to achieve. However, a more in-depth analysis should be performed using additional music material and multiple analysts for increased validity.

The results of these experiments show how vastly different the requirements for contemporary music performance can be from classical music. The main difference lies in the complexity of the communication required. Additionally, this research shows how telepresence technology can benefit contemporary performance practice, since some parameter values are preferred that can only be created within a telepresence environment. It also relates to current research in semio-acoustics and auditory scene analysis [4,30,44] by showing that individuals use different parameters to interpret the same situation, but that certain parameters take precedence for all individuals, that studies of complex auditory scenes may be possible in the laboratory (that indeterminate music is a good example of such a scene), and that meaning-making may be more ambiguous in situations of complex communication. Finally, this research makes an attempt to add to the growing phenomenological library of modes of perception/interaction in which acoustics plays a crucial role, and challenges the idea that one set of acoustic requirements is valid for all types of musical interaction. Husserl's words aptly describe the importance of taking into account each individual and each performance practice when determining the acoustic requirements for a space, and also of updating these requirements constantly.

> "Everything that exists...refers us to evidence, not however to a particular evidence as a de facto experience, but rather to certain potentialities...of the infinity of intendings of every kind...also to the potentiality of verifying these intendings, consequently to potential evidences which, as de facto experiences, are repeatable *in infinitum*" [28].

References

1. Barbour, J. "The Three Dimensional Immersive Soundfield: Delivering Reality." *Proceedings of the Society of Motion Picture and Television Engineers Conference* (Sydney, 2001).

2. Beranek, L. *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*. 2nd ed. New York: Springer, 2004.

3. Beranek, L. *Music, Acoustics and Architecture.* New York: John Wiley & Sons, 1962.

4. Blauert, J. "Analysis and Synthesis of Auditory Scenes." *Communication Acoustics*. Ed. J. Blauert. Berlin: Springer, 2005. 1-25.

5. Braasch, J. *The telematic apparatus – seen from an instrument builder perspective.* Courtesy of the author.

6. Braasch, J., Valente, D., and Peters, N. "Sharing Acoustic Spaces over Telepresence using Virtual Microphone Control." *Proceedings of the Audio Engineering Society Convention* (New York, 2007).

7. Braasch, J., and Woszczyk, W. "A 'Tonmeister' Approach to the Positioning of Sound Sources in a Multichannel Audio System." *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, 2005).

8. Bregman, A.S. *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge, MA: MIT Press, 1990.

9. Brungart, D.S., Kordik, A.J., and Simpson, B.D. "Audio and Visual Cues in a Two-Talker Divided Attention Speech-Monitoring Task." *Human Factors* 47 (2005): 562–573.

10. Brungart, D.S. and Simpson, B.D. "The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal." *Journal of the Acoustical Society of America* 112 (2002): 664-676.

11. Brungart, D.S. and Simpson, B.D. "Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty." *Journal of the Acoustical Society of America* 115 (2004): 301-310.

12. Chew, E., Sawchuk, A., Kyriakakis, C., Papadopoulos, C., François, A., Kim, G., Rizzo, A. and Volk, A. "Musical Interaction at a Distance: Distributed Immersive Performance." *University of Southern California Viterbi School of Engineering, Integrated Media Systems Center.*

13. Cooperstock, J. "Interacting in Shared Reality." *Proceedings of the Human-Computer Interaction International Conference* (Las Vegas, 2005).

14. Cusack, R. and Carlyon, R. P. "Auditory Perceptual Organization Inside and Outside the Laboratory." *Ecological Psychoacoustics*. Ed. J.G. Neuhoff. San Diego: Elsevier, 2004. 16-48.

15. Dammerud, J., and Barron, M. *Stage Acoustics in Concert Halls.* University of Bath Transfer Report (2006).

16. Durlach, N.I. et al. "Informational Masking for Simultaneous Nonspeech Stimuli: Psychometric Function for Fixed and Randomly Mixed Maskers." *Journal of the Acoustical Society of America* 118 (2005): 2482–2497.

17. Fastl, H. "Psycho-Acoustics and Sound Quality." *Communication Acoustics*. Ed. J. Blauert. Berlin: Springer, 2005. 139-162.

18. Francois, A., Chew, E., and Thurmond, D. "Visual Feedback in Performer-Machine Interaction for Musical Improvisation." *Proceedings of the New Interfaces for Musical Expression Conference* (New York, 2007).

19. Gade, A.C. "Musicians ideas about room acoustic qualities." *Technical University of Denmark Report No. 31* (1981).

20. Gade, A.C. "Subjective Room Acoustic Experiments with Musicians." *Technical University of Denmark Report No. 32* (1982).

21. Gade, A.C. "Acoustics of the Orchestra Platform from the Musician's Point Of View." *Acoustics for Choir and Orchestra.* Ed. S. Ternström. Royal Swedish Academy of Music 52 (1986): 23-42.

22. Gade, A.C. "Investigations of Musicians' Room Acoustic Conditions in Concert Halls. Part 1: Methods and Laboratory Experiments." *Acustica* 69 (1989): 193-203.

23. Gade, A.C. "Investigations of Musicians' Room Acoustic Conditions in Concert Halls. Part II: Field Experiments and Synthesis of Results." *Acustica* 69 (1989): 249-262.

24. Garau, M., Friedman, D., Widenfeld, H., Antley, A., Brogni, A., Slater, M. "Temporal and Spatial Variations in Presence: Qualitative Analysis of Interviews from an Experiment on Breaks in Presence." *Presence* 17 (2008): 293-309.

25. Grey, J.M. "Multidimensional perceptual scaling of musical timbres." *Journal of the Acoustical Society of America* 61 (1977): 1270-1277.

26. Heiman, G.W. *Understanding Research Methods and Statistics: An Integrated Introduction for Psychology.* 2nd ed. Boston: Houghton Mifflin, 2001.

27. Henderson, P. "Achieving Perceptually-Accurate Aural Telepresence" (PhD Dissertation, Rensselaer Polytechnic Institute, 2006).

28. Husserl, E. *Cartesian Meditations*: *An Introduction to Phenomenology*. The Hague: Martinus Nijhoff, 1964.

29. ISO-3382. "Acoustics – Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters," *2nd ed.* (1997).

30. Jekosch, U. "Assigning Meaning to Sounds – Semiotics in the Context of Product-Sound Design." *Communication Acoustics*. Ed. J. Blauert. Berlin: Springer, 2005. 193-221.

31. Jones, M.R. "Attention and Timing." *Ecological Psychoacoustics*. Ed. J.G. Neuhoff. San Diego: Elsevier, 2004. 49-88.

32. Jordan, V.L. "Einige Bemerkungen uber Anhall und Anfangsnachhall in Musikraumen." *Applied Acoustics* 1 (1968): 29-36.

33. Kidd, G. Jr., Mason, C., Rohtla, T., and Deliwala, P. "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns." *Journal of the Acoustical Society of America* 104 (1998): 422-431.

34. Kidd, G. Jr., Mason, C.R., and Arbogast, T.L. "Informational Masking Caused by Contralateral Stimulation." *Journal of the Acoustical Society of America* 113 (2003): 1594-1603.

35. Kidd, G. Jr., Mason, C.R., Brughera, A., and Chiu, C. "Discriminating harmonicity." *Journal of the Acoustical Society of America* 114 (2003): 967-977.

36. Kirkwood, B. "Audibility of Changes in Source Directivity for Room Acoustic Auralizations." *Ørsted-DTU Acoustic Technology* (2003).

37. Kohlrausch, A. and Van de Par, S. "Audio-Visual Interaction in the Context of Multi-Media Applications." *Communication Acoustics*. Ed. J. Blauert. Berlin: Springer, 2005. 109-138.

38. "Kolmogorov-Smirnov Goodness-of-Fit Test." *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm, 7/18/2006.

39. Krokstad, A., Vindspoll, J., Saether, R. *Orkesterpodium, Samspill of Solo*. Laboratory of Acoustics, the University of Trondheim (1980).

40. Marshall, A.H., Gottlob, D., Alrutz, H. "Acoustical Conditions Preferred for Ensemble." *Journal of the Acoustical Society of America* 64 (1978): 1437-1442.

41. Meyer, J. *Acoustics and The Performance of Music.* Frankfurt am Main: Verlag das Musikinstrument, 1978.

42. Meyer, J. and Biassoni de Serra, E.C. "Verderckungseffekt bei Instrumentalmusikern." *Acustica* 46 (1980): 130-140.

43. Möller, S. "Quality of Transmitted Speech for Humans and Machines." *Communication Acoustics*. Ed. J. Blauert. Berlin: Springer, 2005. 163-192.

44. Neuhoff, J.G. "Ecological Psychoacoustics: Introduction and History." *Ecological Psychoacoustics*. Ed. J.G. Neuhoff. San Diego: Elsevier, 2004. 1-13.

45. Novo, P. "Auditory Virtual Environments." *Communication Acoustics*. Ed. J. Blauert. Berlin: Springer, 2005. 277-297.

46. Nyman, M. *Experimental Music: Cage and Beyond*. 2nd ed. Cambridge: Cambridge University Press, 1999.

47. Oliveros, P. "Acoustic and Virtual Space as a Dynamic Element of Music." *Leonardo Music Journal* 5 (1995): 19-22.

48. Oliveros, P. "Improvising with Spaces." *Proceedings of the 13th International Conference on Auditory Display* (Montreal, 2007).

49. Otondo, F., Rindel J.H., Caussé, R., Misdariis, N. and Cuadra, P. "Directivity of musical instruments in a real performance situation." *Ørsted-DTU Acoustic Technology.*

50. Oxenham, A., Fligor, B., Mason, C.R., and Kidd, G. Jr. "Informational masking and musical training." *Journal of the Acoustical Society of America* 114 (2003): 1543-1549.

51. Parsons, M. "The Scratch Orchestra and Visual Arts." *Leonardo Music Journal* 11 (2001): 5-11.

52. Peruch, P., Belingard, L., Thinus-Blanc, C. "Transfer of Spatial Knowledge from Virtual to Real Environments." *Spatial Cognition II*. Ed. C. Freksa, et al. Berlin: Springer, 2000. 253-264.

53. Richards, V.M. and Neff, D.L. "Cuing effects for informational masking." *Journal of the Acoustical Society of America* 115 (2004): 289-300.

54. Rosenblum, L.D. "Perceiving Articulatory Events: Lessons for an Ecological Psychoacoustics." *Ecological Psychoacoustics*. Ed. J.G. Neuhoff. San Diego: Elsevier, 2004. 220-249.

55. Schroeder, R., et al. "Collaborating in networked immersive spaces: as good as being there together?" *Computers & Graphics* 25 (2001): 781–788.

56. Shinn-Cunningham, B. and Streeter, T. "Perceptual Plasticity in Spatial Auditory Displays." *Proceedings of the International Conference on Auditory Display* (Helsinki, 2001).

57. Shinn-Cunningham, B. and Ihlefeld, A. "Selective and Divided Attention: Extracting Information from Simultaneous Sound Sources. *Proceedings of the International Conference on Auditory Display* (Sydney, 2004).

58. Proceedings of the SIGGRAPH 2007 conference, CALIT Auditorium, University of California, San Diego with two remote locations at RPI and CCRMA Stanford (Aug. 5) and Banff Centre (Aug. 6), Aug 5 and 6, 2007.

59. Ternström, S. and Sundberg, J. "Acoustics of Choir Singing." *Acoustics for Choir and Orchestra.* Ed. S. Ternström. Royal Swedish Academy of Music 52 (1986): 23-42.

60. Ternström, S., Cabrera, D., and Davis, P. "Self-to-other ratios measured in an opera chorus in performance." *Journal of the Acoustical Society of America* 118 (2005): 3903-3911.

61. Thomasson, S.E. "Kan god podieakustik matas?" Proceedings of the Scandinavian Acoustical Society (1974).

62. Ueno, K. and Tachibana, H. "Experimental Study on the Evaluation of Stage Acoustics by Musicians Using a 6-Channel Sound Simulation System." *Acoust. Sci. & Tech.* 24 (2003): 130-138.

63. Ueno, K., Kanamori, T. and Tachibana, H. "Experimental study on stage acoustics for ensemble performance in chamber music." *Acoust. Sci. & Tech*. 26 (2005): 345-352.

64. Ueno, K., Kato, K., and Kawai, K. "Musicians Adjustment of Performance to Room Acoustics, Parts I, II, and III." *Proceedings of the International Congress on Acoustics* (Madrid, 2007).

65. Valente, D. and Braasch, J. "Perceptual adjustment of direct-to-reverberant ratio and reverberation time to match visual environmental cues of a musical performance." *Acustica* (2008): Special Issue on Virtual Acoustics (under review).

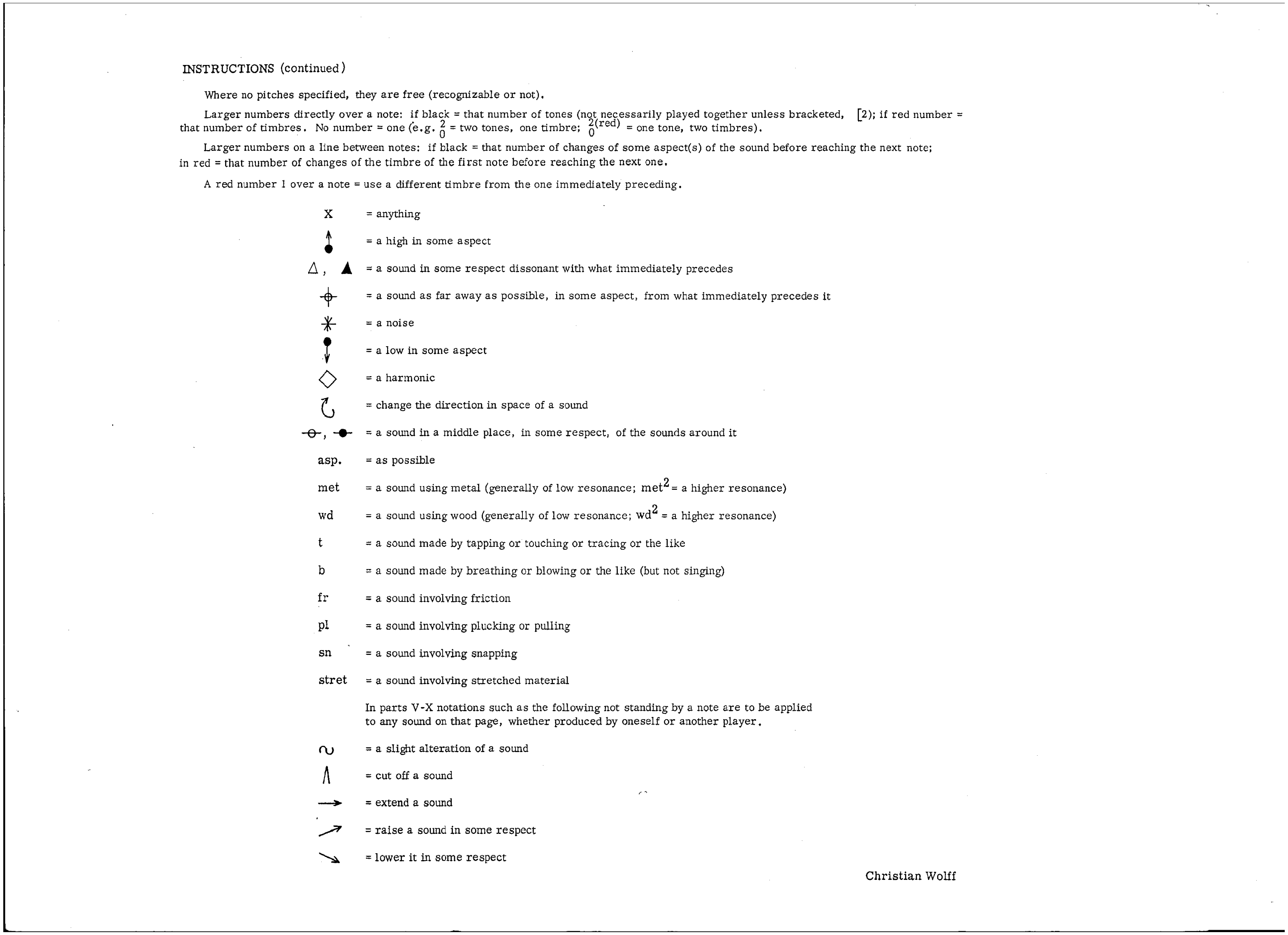
66. Van Valkenburg, D. and Kubovy, M. "From Gibson's Fire to Gestalts: A Bridge-Building Theory of Perceptual Objecthood." *Ecological Psychoacoustics*. Ed. J.G. Neuhoff. San Diego: Elsevier, 2004. 114-149.

67. Walker, B.N. and Kramer, G. "Ecological Psychoacoustics and Auditory Displays: Hearing, Grouping, and Meaning-Making." *Ecological Psychoacoustics*. Ed. J.G. Neuhoff. San Diego: Elsevier, 2004. 150-175.

68. Wolff, Christian. *For 1, 2 or 3 People.* New York: C.F. Peters, 1964.

69. Zha, X., Fuchs, H.V., Drotleff, H. "Improving the Acoustic Working Conditions for Musicians in Small Spaces." *Applied Acoustics* 63 (2002): 203-221.

Appendix



INSTRUCTIONS (continued)

Where no pitches specified, they are free (recognizable or not).

Larger numbers directly over a note: if black = that number of tones (not necessarily played together unless bracketed, [2); if red number = that number of timbres. No number = one (e.g. $\frac{2}{0}$ = two tones, one timbre; $\frac{2(\text{red})}{0}$ = one tone, two timbres).

Larger numbers on a line between notes: if black = that number of changes of some aspect(s) of the sound before reaching the next note; in red = that number of changes of the timbre of the first note before reaching the next one.

A red number 1 over a note = use a different timbre from the one immediately preceding.

X = anything

= a high in some aspect

△, ▲ = a sound in some respect dissonant with what immediately precedes

= a sound as far away as possible, in some aspect, from what immediately precedes it

= a noise

= a low in some aspect

◇ = a harmonic

= change the direction in space of a sound

= a sound in a middle place, in some respect, of the sounds around it

asp. = as possible

met = a sound using metal (generally of low resonance; met^2 = a higher resonance)

wd = a sound using wood (generally of low resonance; wd^2 = a higher resonance)

t = a sound made by tapping or touching or tracing or the like

b = a sound made by breathing or blowing or the like (but not singing)

fr = a sound involving friction

pl = a sound involving plucking or pulling

sn = a sound involving snapping

stret = a sound involving stretched material

In parts V-X notations such as the following not standing by a note are to be applied to any sound on that page, whether produced by oneself or another player.

∾ = a slight alteration of a sound

= cut off a sound

→ = extend a sound

↗ = raise a sound in some respect

↘ = lower it in some respect

Christian Wolff

Figure 27: Christian Wolff: *For 1, 2 or 3 People*, Instructions Pg. 1 [68].

INSTRUCTIONS

There are ten parts, one to a page. A performance can be made of any number of them, repeating none, or of any one, repeated no more than ten times.

Each part, or page, is a score, and each player should have his copy of it.

Play all that is notated on a page, in any convenient sequence, not repeating anything; except in IX, where any of the events can be played or omitted any number of times.

Black notes are variously short, up to about one second. With stems as sixteenth notes (e.g. in III, etc.) they are very short. White notes are of any length, sometimes determined by the requirements of coordination (see further on).

A dynamic indication may stand by itself (as at left top of I): assume a note to go with it or apply it to any note given on the page. However, or , standing by themselves, should always be applied to a note (any one) already given.

A diagonal line towards a note = play that note directly after a preceding one. A diagonal line away from a note = that note must be followed directly by another.

A vertical line down from a note = play simultaneously with the next sound (both attack and release).

A small number at the end of a line (e.g. at left top of I) = coordinate with the second (if the number is 2; third, if 3; etc.) sound, preceding (if diagonal line towards note), following after one has begun one's note (if diagonal line away from it), or play simultaneously with the second next sound (if the line is vertical).

If a line to a note is broken by a number followed, after a colon, by a zero (——2:0——) (e.g. top middle of III), that number of seconds of silence intervene before the required coordination.

An at the end of a line (e.g. middle left in I) = coordination must be with a sound made by another player. If only one person is playing, he must coordinate either with a sound he hears in the environment or with a sound he has himself made unintentionally.

= play after a previous sound has begun, hold till it stops.

= start anytime, hold till another sound starts, finish with it.

= start at the same time (or as soon as you are aware of it) as the next sound, but stop before it does.

= start anytime, hold till another sound starts, continue holding anytime after that sound has stopped.

Horizontal lines joining two notes = a legato from the one to the other (both played by the same person).

If no line leads to a note or drops vertically from it, one can start to play at any time. If no line leads away from a white note, it can last as long or as short as you like.

One, two, or three people can play. If one plays alone, he must realize all "open" coordinations (lines with notes at only one end) himself, that is, he must use other notes given on a page, as he can to provide something to coordinate with; or, sometimes, he may use sounds from the environment (as he must when there is an at the end of a line). (He may in some cases have to rearrange the material on a page and consider a disposition of it which will ensure that all the required coordinations can be managed.) All the material on a page can be freely superimposed, so long as the requirements of coordination are met.

If two or three play, the material on a page should be distributed between them, in any way (in VII a distribution for two players is indicated); but no material marked off for one player should be played by another (note: this holds for IX too). Coordination, then, for each player can be either with his own material (as if he were playing alone)--unless there is an -- or with whatever sound(s) he hears next from another player (or both).

Players can use any ways of making sounds, allowing for the following specifications:

Some notes are on staves: play the indicated pitch (reading either bass or treble clef; sound at pitch; if pitch not available in range, transpose at least two octaves; short lines off a pitch at an angle = fraction of a tone less than half up where line angles up, down where down).

Figure 28: Christian Wolff: *For 1, 2 or 3 People*. Instructions Pg. 2 [68].

Command	Parameter(s)	Description
/SourceXpos	index m, x [m]	changes x position for sound source with index m
/SourceYpos	index m, y [m]	changes y position for sound source with index m
/SourceZpos	index m, z [m]	changes z position for sound source with index m
/SourcePos	index m, x [m], y [m], z [m]	changes position (x,y,z) for sound source with index m
/RoomSize	x [m], y [m], z [m]	changes the room size of the virtual room (x,y,z)
/RoomWidth	x [m]	changes width x of virtual room
/RoomDepth	y [m]	changes depth y of virtual room
/RoomHeight	z [m]	changes height z of virtual room
/MicXpos	index n, x [m] or only x [m] to address all mics	changes x position of virtual microphone with index n
/MicYpos	index n, y [m] or only y [m] to address all mics	changes y position of virtual microphone with index n
/MicZpos	index n, z [m] or only z [m] to address all mics	changes z position of virtual microphone with index n
/MicPos	index n, x [m], y [m], z [m]	changes position (x,y,z) of virtual microphone with index n
/MicCenterDistance	d [m]	positions all microphones at distance d from the center of the microphone array at their current angles
/MicAzi	index n, *alpha* [deg] or only azi to address all mics	determines azimuth angle *alpha* for the directivity pattern of the virtual mic
/MicAzi	index n, *theta* [deg] or only azi to address all mics	determines elevation angle *theta* for the directivity pattern of the virtual mic
/MicAngle	index, *alpha* [deg], *theta* [deg]	determines azimuth and elevation angle of virtual mic n
/Directivity	Γ	determines the directivity pattern of a mic with 0= index n, figure-8, 0.5=cardioid, 1=omni
/DirPow	index n, δ	provides the directivity power for mic n $\delta = 1$ 1st-order microphone
/DisPow	index n, r	distance power, determines the amplitude decay with distance, 1=1/r law, 0=no amplitude decay with distance
/ReportAll	bang	ViMiC will print out the following data: number of channels, source positions, room size (x,y,z), Mic Array Center (x,y,z) and microphone data including delay and sensitivity
/Report	1 (on), 0 (off)	will print out every executed command with variables if /Report is set to 1

Table 1: Implemented OpenSound Control commands for ViMiC.

Figure 29: ViMiC OSC Commands [6].

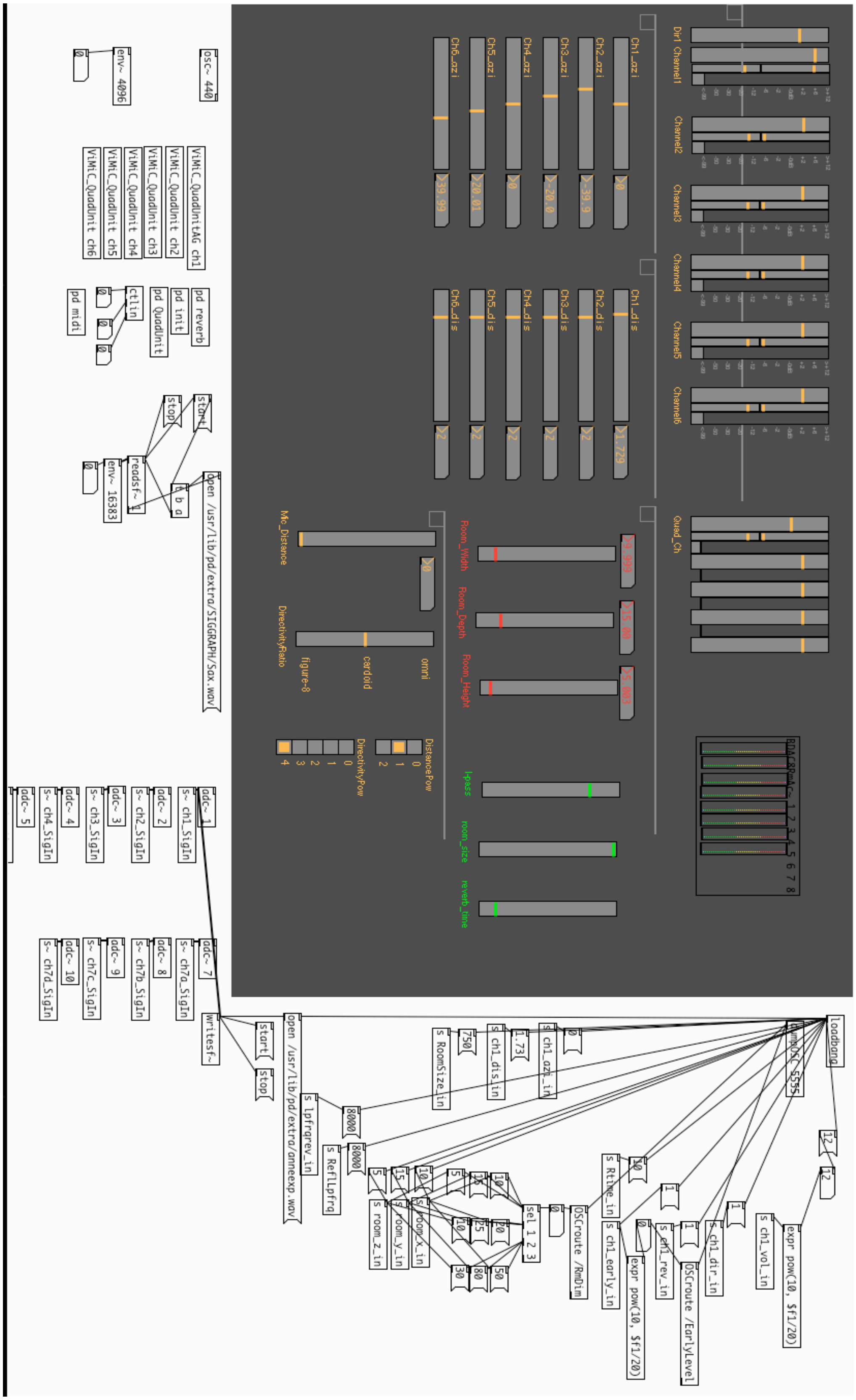


Figure 30: Sample ViMiC Patch (Early Reflected Energy)

```
Room A:
echo Transmission start up file 2/2
killall qjackctl
killall jackd
killall jacktrip
killall pd
read -p "Press any key to start Pure data"
pd -verbose -lib Gem -lib oscx -jack -channels 12 -realtime –path
/usr/lib/pd/extra/SIGGRAPH -open VimicGUILEVELRmA.pd &
read -p "Press any key to start qjackctl"
qjackctl &
read -p "Press any key to start Jacktrip server Fenster"
echo Connection to Fenster
jacktrip -j -z 64 -b 4 -q 4 -a 4 -t 128.113.106.148 &
jmess -c AnneRouting2.xml


Room B:
echo Transmission start up file 1/2
killall qjackctl
killall jackd
killall jacktrip
killall pd
read -p "Press any key to start Pure data"
pd -verbose -lib Gem -lib oscx -jack -channels 12 -realtime -path
/usr/lib/pd/extra/SIGGRAPH -open VimicGUILEVELRmB.pd &
read -p "Press any key to start qjackctl"
qjackctl &
read -p "Press any key to start Jacktrip server Fenster"
echo Connection to Fenster
jacktrip -j -z 64 -b 4 -q 4 -a 4 -r &
jmess -c AnneRouting1.xml
```

Figure 31: ***Jacktrip*** **Connection Protocol**

```
Room A:
cd /home/jer/uv-2.0-dynamic/
./uv8020 -x –D
./uv8020 server started on port 8020
./uv8020 -x 128.113.106.148

Room B:
cd /home/bronto/uv-2.0-dynamic/
./uv8020 -x –D
./uv8020 -x 128.113.106.136
```

Figure 32: ***McGill Ultra Videoconferencing System*** **Connection Protocol**

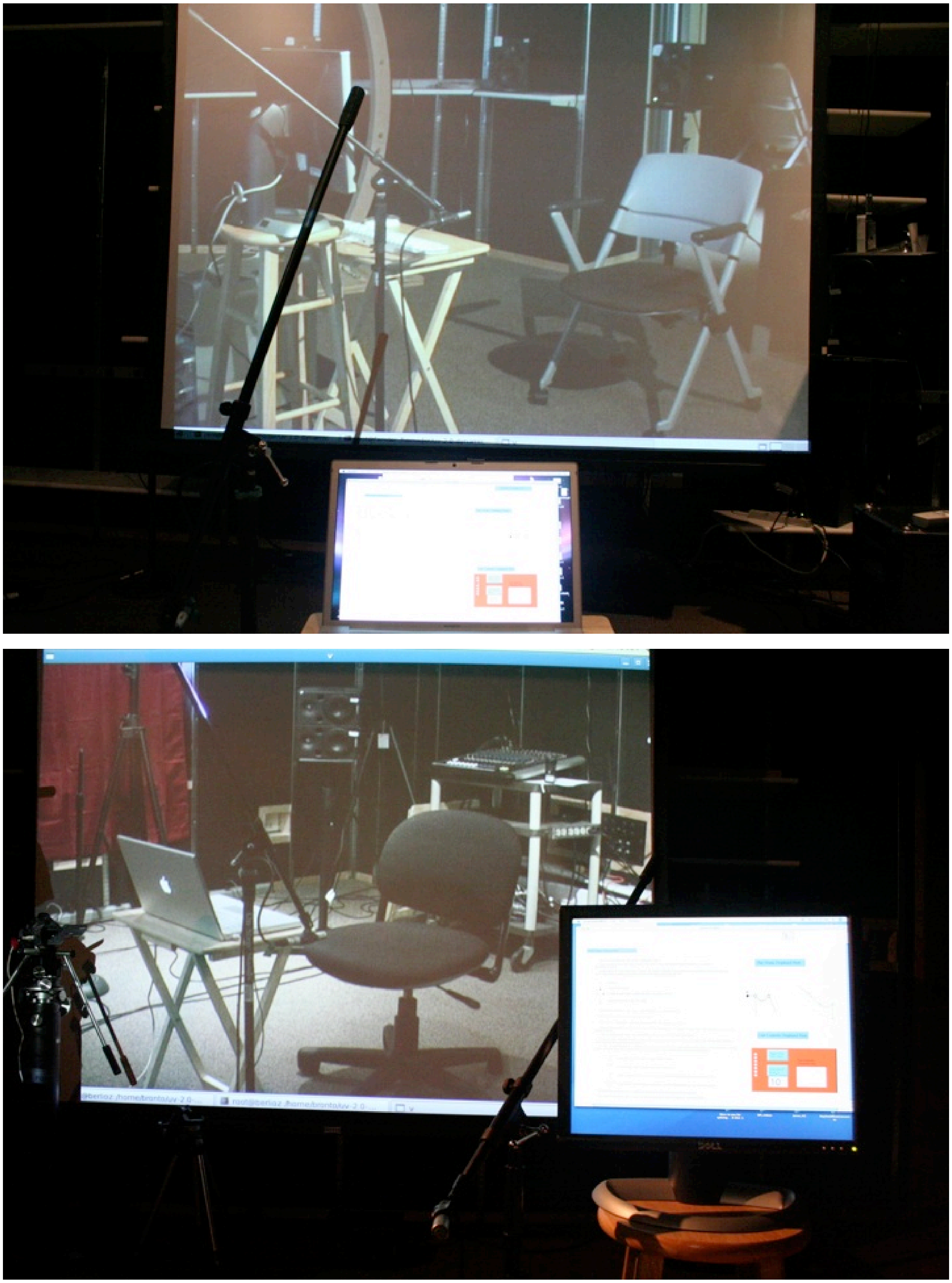


Figure 33: Photographs of Experiment Setup (no Subjects), Rooms A and B, respectively.

Figure 34: Photographs of Experiment Setup: Rooms A and B.

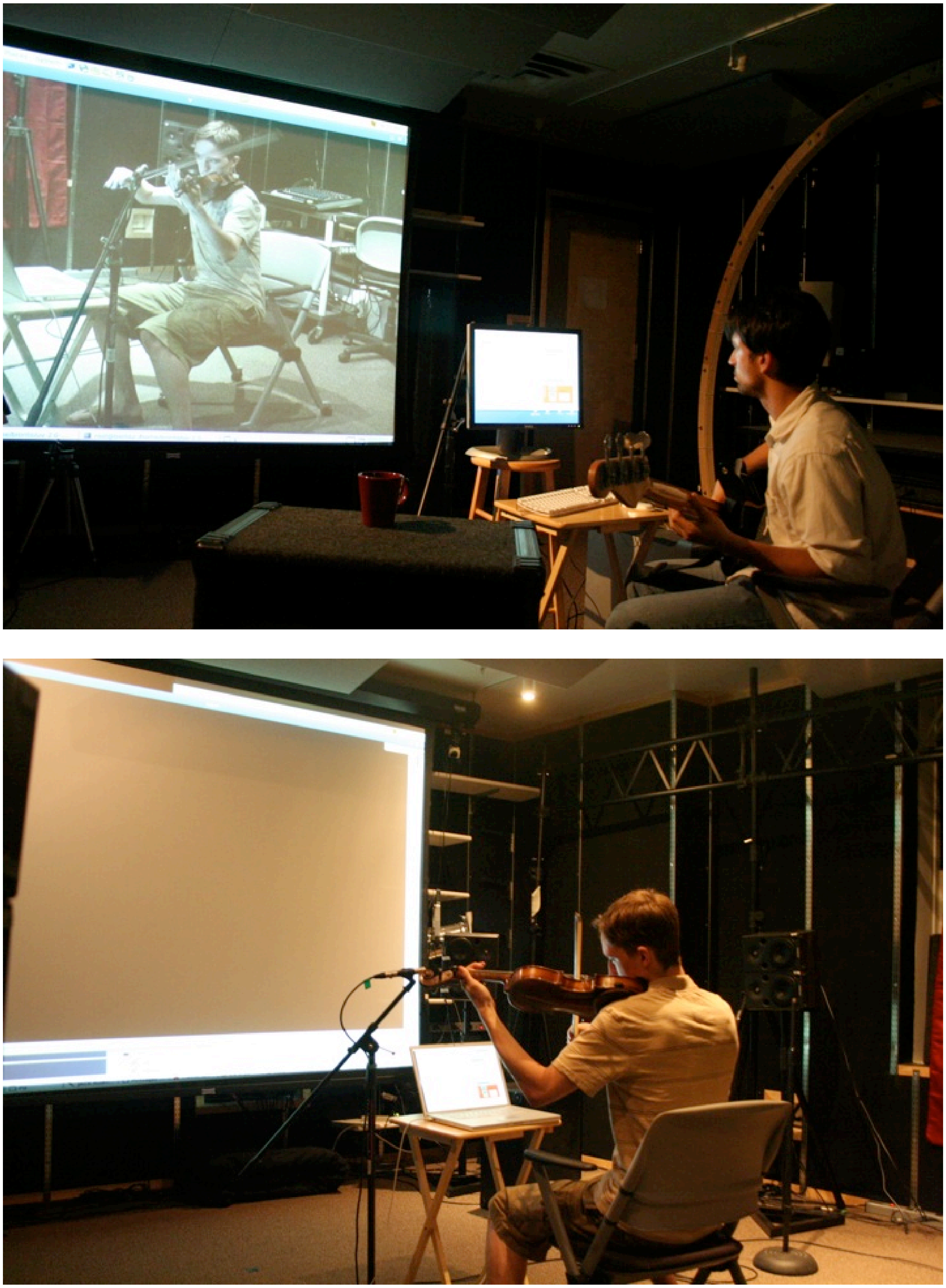


Figure 35: Photographs of Experiment Setup: Rooms B (test with Video) and A (test without Video), respectively.