ICMC 2006

MULTIDIMENSIONALITY

THE INTERNATIONAL COMPUTER MUSIC CONFERENCE
AN ICMC-SEAMUS COLLABORATION

NOVEMBER 6-11, 2006
TULANE UNIVERSITY
NEW ORLEANS, LA USA

www.icmc2006.org

4:30 PM — Real-Time Synchronization of Independently-Controlled Phasors
Lonce Wyse

5:00 PM — A Paradigm For Physical Interaction With Sound In 3-D Audio Space
Mike Wozniewski, Zack Settel, Jeremy Cooperstock

5:30 PM — Jam'aa - A Middle Eastern Percussion Ensemble for Human and Robotic Players
Gil Weinberg, Scott Driscoll, Travis Thatcher

Paper Session 8 B — Diboll Conference Center Room B

Music Analysis

3:30 PM — Recording Quality Ratings by Music Professionals
Richard Repp

4:00 PM — Data Association Techniques for a Robust Partial Tracker of Music Signals
Hamid Satar-Boroujeni, Bahram Shafai, Patric J. Wolfe

4:30 PM — Musical Tension Curves and Its Applications
Min-Joon Woo, In-Kwon Lee

5:00 PM — Detecting Motives and Recurring Patterns in Polyphonic Music
Paul Utgoff, Phillip Kirlin

5:30 PM — Melodic Modeling: A Comparison of Scale Degree and Interval
Yipeng Li, David Huron

Recording Quality Ratings by Music Professionals

Richard Repp, Ph.D.
Department of Music, Georgia Southern University
rrepp@richardrepp.com

Abstract

This study explored whether music professionals can perceive quality differences in recordings of classical musicians on acoustic instruments. Thirty-two music professionals listened to a series of twelve recordings at nine differing quality levels. Quality levels included pristine 24 bit, 192 kHz recordings, Compact Disk (CD) quality recordings, cassette tapes, MP3 files, and recordings with noise added. The participants judged the quality of the recordings. A one-way ANOVA test found significant differences among the responses from groups (F=302, p<.001), and a Tukey HSD test determined which groups were significantly different. The study proved that music professionals are able to hear the difference between a CD quality recording and the same recording transferred to cassette tape. For this reason, the researcher recommends using a CD recording for audition purposes rather than a cassette tape. However, extremely high quality recordings above standard CD quality received equivalent ratings to CD recordings.

1. Introduction

Emerging technologies have made high fidelity recording a possibility. In today's market, musicians are often judged by recorded material they produce rather than a live, acoustical presentation. In audition situations, musicians are often asked to produce a recording in lieu of expensive visits to locations. Unfortunately, not all musicians have access to high-end equipment to produce quality recordings.

Some musicians do have the opportunity to record in a professional recording studio, but the cost of such recordings can be prohibitive. Furthermore, research has not established whether the quality of such recordings is noticeable to audition judges. This study provides an answer to the question of whether recording quality differences are noticeable to music professionals.

1.1 Background

Use of digital audio for the dissemination of music has captured the attention of the national media with recent court cases involving Napster and other file-sharing systems. Digital listening is now commonplace with hardware ranging from the compact disk to the MP3 player to high-fidelity DVD-Audio formats. What remains to be seen is whether professional musicians require the higher level of fidelity to judge recorded musical performance. The difference is nontrivial when hiring or acceptance to an institution of higher learning is at stake.

The field of music is highly competitive, and even a small advantage in perceived skill could make the difference in an audition. However, if the added expense of high-end recording does not result in higher scores at an audition or competition, then valuable resources have been wasted.

1.2 Need

Although digital music has received much attention in the popular music realm, the use of recording for classical musicians has less of a research base. Most attention to the use of technology in classical or academic music goes into composition or music education. Since performance is such an integral part of what musicians do, research into the use of technology lags behind the actual needs of the profession.

One meaningful use of technology for performance professionals and students is the use of recordings for archival and review purposes. Use of recordings for initial auditions is commonplace both in academia and industry, as noted in the following audition requirements:

> "Please send your resume to the Orchestra including the information below and also a CD/ MD/Video/DVD recording of a performance you have given in the last 2 years." Tokyo Philharmonic Orchestra.
>
> "Tapes may be either digital, analog, or VHS, but the microphone(s) used should be

of sufficient quality to provide an accurate picture of your work." Wittenberg College.

"If you cannot arrange an in-person audition, you may submit a high quality audio cassette or CD recording." Northwestern University.

"In addition to the video audition you are welcome to send additional material—CD, cassette or video of live performances, studio or home recordings, lyric sheets, bios or reviews." University of Otago.

"You can audition in person … or you can send a CD, tape or even video. Make this as high quality as possible." St. Francis Xavier University.

"Applicants from outside the United States may send a CD of the required audition materials. Any evidence of tampering of the recording will disqualify the applicant." Civic Orchestra of Chicago.

Many of the audition announcements mention the importance of high quality recordings, but none specifically define what an acceptable level of quality is.

2 Research Literature

Although research directly applicable to music auditions is extremely limited, a wealth of information on the recording process exists. Most applicable to the present research include those studies on the recording environment (McKinnie, 1991; 1996; Møller, Sørensen, Jensen, & Hammershøi, 1996; Newell & Holland, 1997). These studies stress the importance of a controlled listening environment and its relationship to the perception of music. Applicable research on the recording process also includes Gabrielsson, Hagreman, Bech-Kristensen, and Lundberg (1990) and Lipshitz (1986). Some research directly addresses the issue of whether the high-frequency possibilities of high-sampling rate recordings actually improves sound quality (e.g., Ohashi, Nishina, Kawai, Fuwamoto, & Imai, 1991; Ohashi, Nishina, Fuwamoto, and Kawai, 1993; Zielinski S.K., Rumsey, & Bech, 2002).

For purposes of designing testing mechanisms, several evaluation scenarios were explored (Bareham, 1996; Bech, 1987; Hansen & Munch, 1991, Meilgaard, Civille, & Carr, 1991), with an emphasis on those systems that test subjective reactions to recordings rather than technical readings (Grewin, 1995; Guski, 1997; Precoda, & Meng, 1997; Stuart, 1991; Toole, 1985). More general work includes studies on perception (Bregman, 1990; International Telecommunications Union, 1997; Griesinger, 1997; 2001; Mason & Rumsey, 2000; Terhardt, 1990; Umemoto, 1990; Rumsey, 1999) and subjectivity (Berg & Rumsey, 2000; Kirk, 1956; Kosslyn, 1981; Meares, 1993; Moore, 1997). Applicable research on acoustics is plentiful (e.g., Ando, 1998; Blauert & Lindemann, 1986; Mapp, 1997).

Some research exists on the relationship between quality of recordings and enjoyment of music. Research indicates that the cost of an audio system does not have a statistical correlation to appreciation of the art. Roy Harris (2002) writes,

> Currently, there is no evidence that music appreciation is dependent on sound quality. This means that one can attain the same level of musical enjoyment from any medium as long as the flaws in the components do not render the sound unpalatable. The reason one enjoys the music when listening uncritically has little to do with the quality of one's stereo system, as the sound quality is not a predictor of the affect music has on a listener.

Mark Sauer (2000) also found that "…greater accuracy does not mean more pleasure. If the sound quality of stereo systems is not a significant contributor to a satisfactory listening experience, what is? The answer may reside within the listener."

However, little hard research exists on the correlation between quality of recorded audio and perception of the performer in an audition situation. In fact, most of the research in the area of auditioning is not experimental, and is more experiential (e.g., Legge, 1990). In a professional environment the listener is less interested in the enjoyment of music, as stressed in the research, and more interested in the skills of the applicant.

3 Methodology

Research Question: Are recording quality differences noticeable to music professionals?

Auditioners are interested in whether a high-quality recording affects their score on auditions. But in order to answer this question, first the level at which potential audition judges can notice quality differences takes precedence.

3.1 Participant Selection

After obtaining permission from an Institutional Review Board, participants (N=32) gave permission to take art in the experiment. All experimental participants were music professionals, mostly university professors. Five of the participants were graduate students who had worked in the past as music professionals. Participants were not selected randomly from a larger population group. Participant selection emphasized real-world experience in auditions so that the results could be generalized to the population of music professionals likely to hear audition recordings.

3.2 Procedures

The experimenters produced recordings of four different instruments—French horn, flute, clarinet, and voice—with three recordings each, for a total 12 separate recordings. All selections were recorded dry (with no reverberation, natural or artificial), with no accompaniment.

All recordings took place in the same room with the same equipment and setup. Two Neumann KM-184s in a stereo configuration recorded through a Mark of the Unicorn (MOTU) 896 analog to digital converter (ADC) into MOTU Digital Performer software. The original bit rate and sample rate of the recordings was at 24 bit, 192 kHz (defined here as very high quality). Normalized recordings (amplified to maximum possible level) assured that judgments were not affected by volume differences.

Then, data reduction procedures reduced the quality of each of the recordings eight times, for a total of 9 data groups. The original recordings of 24 bit, 192 kHz went through a translation to 16 bit 44.1 kHz (standard CD quality). A third group was in the popular MP3 format at the standard 128 kbps data rate. The third group represented medium fidelity in today's digital world. The fourth data group consisted of the original examples recorded to cassette tape.

Additional groups included the original recording mixed with differing levels of pink noise. The reference value for mixing of pink noise would be from a level of "0" having equal amounts of pink noise as the original signal; the next highest quality signal (presumably) was –60 dB pink noise (60 dB softer than equal amounts). Then groups of –50 dB, -40 dB, -30 dB, and –15 dB added pink noise completed the nine groups.

The total number of samples, 12 recordings at nine quality levels for a total of 108 examples, was too large, so a stratified sample provided a final grouping. Three examples of each of the original recordings were chosen at three different levels, so that each of the nine quality groups had four samples, for a total of 36 items in the final data list. All musical examples, quality examples, and instrument groups had an equal number of items in the final set. The final set contained the 36 examples put into a random order using a software-driven randomizer.

3.3 Data Collection

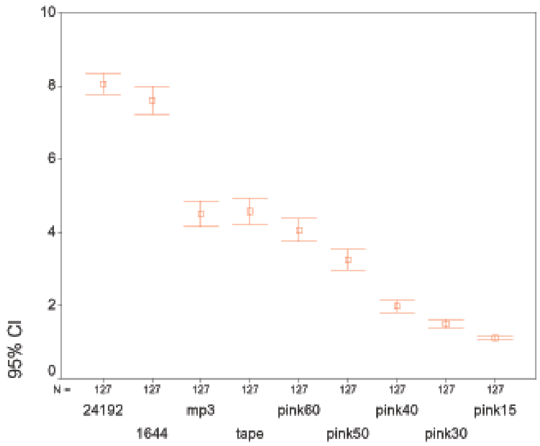
The participants listened to the examples in a quiet (less than 30 dB SPL ambient noise), acoustically balanced room. Monitors (speakers) consisted of Tannoy Reveal Monitors placed one meter from the subject at the corners of an imaginary equilateral triangle. All participants sat in the same chair, which was in the same position (the third corner of the triangle), for every session. Before the session began, the experiments tested the audio to confirm that the volume levels were consistent (~78 dB SPL) using and SPL meter.

Before the experiment, a recorded voice reminded the participants that they were judging the quality of the recording, and not the performance of the person recorded. The participants rated the recording quality on a ten point Likert-type scale, with 10 being the best possible recording. Testers did not coach the participants as to what "good" or "bad" quality was. If the participants asked questions concerning the definition of quality before the experiment began, they were told to use their best judgment.

3.4 Results

Figure 1 shows the relative scores for the means each of the quality comparison groups with their 95% confidence interval.

Figure 1. Error Plot of Relative Means of Scores for Quality Groups (95% CI).



A One-way ANOVA test using SPSS software showed that there is significant differences among the groups at p<.001, with F=301.6. (See Figure 2).

Figure 2. ANOVA Results.

Test of Homogeneity of Variances

fulldata

Levene Statistic	df1	df2	Sig.
45.837	8	1142	.000

ANOVA

fulldata

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6246.318	8	780.790	301.693	.000
Within Groups	2955.531	1142	2.588		
Total	9201.849	1150			

Post hoc tests (Tukey HSD) showed that all groups were significantly different from each other (p<.002) *except* the following pairs (See Figure 3):

Figure 3. Tukey HSD Non-significant Differences.

24-192 and 16-44	p = .336
MP3 and Tape	p = 1.000
MP3 and -60 dB Pink	p = .453
Tape and -60dB Pink	p = .336
-40 dB Pink and -30 dB Pink	p = .270
-30 dB Pink and -15 dB Pink	p = .661

Tukey HSD analysis confirmed that homogeneous groups existed (See Figure 4).

Figure 4. Tukey HSD Homogenous Groups.

Tukey HSD[a,b]

Category	N	Subset for alpha = .05				
		1	2	3	4	5
-15bB Pink	128	1.1172				
-30bB Pink	128	1.5000	1.5000			
-40bB Pink	128		1.9883			
-50bB Pink	128			3.2500		
-60bB Pink	127				4.0630	
mp3	128				4.4922	
tape	128				4.5781	
16-44	128					7.5742
24-192	128					8.0391
Sig.		.612	.270	1.000	.205	.336

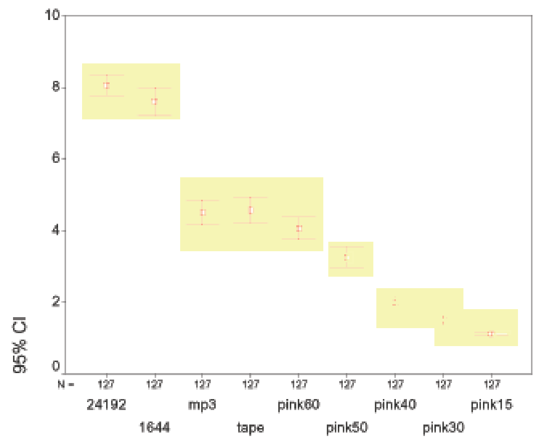
Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 127.888.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

The following graph (Figure 5) shows a graphical representation of the data in Figure 4, with homogeneous subsets connected by shaded areas over the error plots from Figure 1.

Figure 5. Homogeneous Groups Graph.



4 Discussion

Analysis of the descriptive data shows that the test was a valid measure of the participants' ability to judge the quality of the recorded audio. The readings on the variable of the amount of pink noise added to the sample (Figure 1, the final 5 columns) show an incremental decrease in the quality score as the amount of pink noise rises. The high F score (F=306) (See Figure 2) shows that the results were indeed significant and reliable.

The most useful data comes from the post-hoc tests. The homogeneity tests show that the quality levels break into distinct categories.

The first of these categories would be considered high fidelity. Recordings at 24 bit, 192 kHz and recordings at 16 bit, 44.1 kHz were statistically indistinguishable from each other. Even though the 24 bit, 92 kHz scored slightly higher than their counterparts, the difference was not statistically significant. To reinforce this lack of differences, relative scores for the four sub-groups of recordings (horn, clarinet, flute, and voice) show that on three of the four subgroups, the 16 bit 44.1 kHz example actually scored slightly higher than the 24 bit, 192 kHz example. (See Figures 6-9.)

Figure 6. Relative Means for Clarinet Examples.

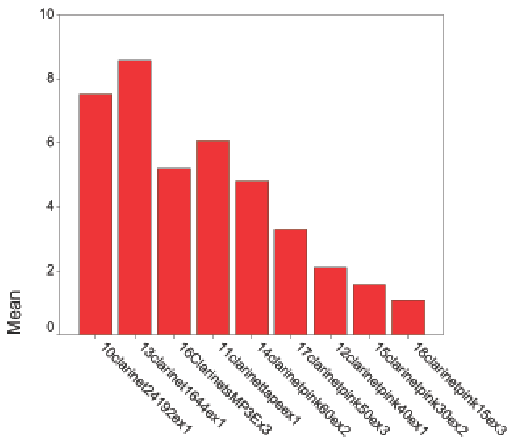


Figure 7. Relative Means for Flute Examples.

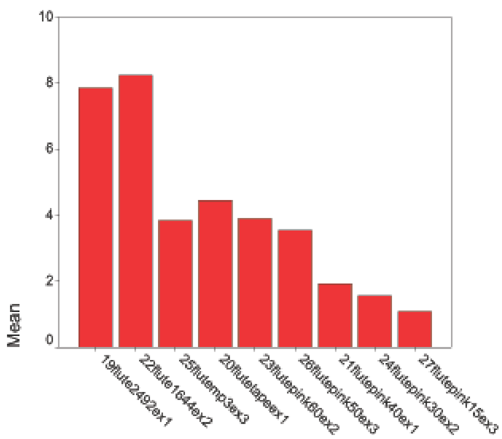
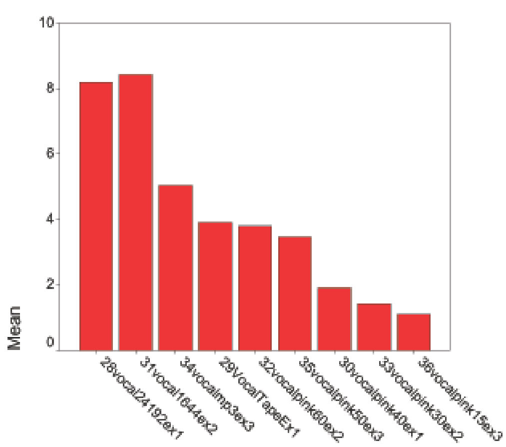
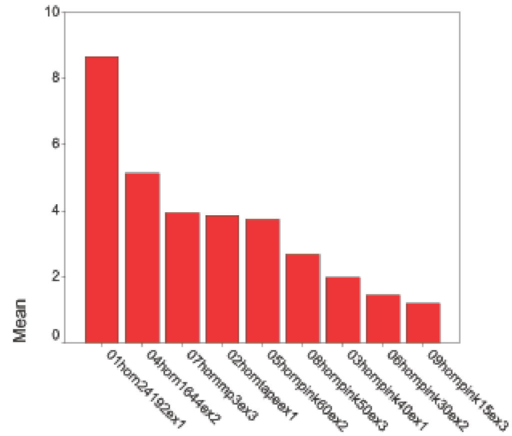


Figure 8. Relative Means for Vocal Examples.



Only the large difference in the horn example (Figure 9) accounts for the final difference.

Figure 9. Relative Means for Horn Examples.



A clear distinction exists between the high fidelity group and the next homogeneous group, which consists of the MP3 sample, cassette tape recordings, and the recording with –60 dB pink noise added (see Figure 5). Music professions were clearly able to hear the difference between a CD quality recording and a cassette quality recording or its equivalent.

One might expect an MP3 recording to sound better than a cassette tape. The lack of difference in these scores could be influenced by several factors. The cassette recordings used in this experiment were of unusually high quality, since they were recordings from a digital source that had been recorded under optimal conditions. The cassettes one might expect to hear in a real-world audition would probably be recorded directly to tape, and presumably would not be as high a quality, even if the same recording setting existed.

The wide variation in possible MP3 qualities could also be a factor. A well-engineered MP3 file is not distinguishable from a CD quality recording. The MP3 files in this experiment were purposely of low quality. Interestingly, the digital artifacts in the MP3 files (jitter) were no more or less distracting to the participants than the inherent noise associated with cassette tape recording.

Readings on the low quality recordings (pink noise added) are less interesting from a real-world perspective because recordings as bad as the worst recordings would never be used in an audition situation. The poor examples were useful in dispersing the Likert-type responses, so that the participants could hear what a truly very bad recording sounds like. The data also shows that the participants were able to distinguish a 10 dB addition of pink noise.

5 Conclusions

Music professionals are able to hear the difference between a compact disk quality recording and the same recording transferred to cassette tape. For this reason, the researcher recommends using a digital CD recording for audition purposes rather than a cassette copy. Music professionals do have a discerning ear for recordings, even though they may have been raised on old, scratched records and hiss-filled tape. However, extremely high quality recordings above standard CD quality are ranked equivalent to CDs by music professionals, so spending the extra money for these recordings is not necessary.

Also, if the music professionals judge the recording quality of 128 kbps MP3 files and cassette tapes equivalent (as shown by this study), this does not mean that this difference in medium will not affect their judgment. The impact on the judgment of the visual quality of the material could also be considered as well as the use of "up-to-date technology". A professional-looking CD-ROM with MP3 files might make a better impression on the judges than an old cassette tape. This should not matter to judge the quality of a performer, but it probably does matter in reality.

Even though this study has proven that musicians can hear these differences, the question still remains as to whether these differences in recording quality lead to improved scores on auditions. Now that the researcher has proven that these differences exist, future studies must prove whether judges ignore the differences, either consciously or unconsciously. Another possibility may be that a poor recording masks flaws in the performance, so that a high-quality recording actually hurts the audition score.

Another question left unanswered is whether music professionals would be able to hear the recording quality differences outside of a controlled listening environment. In order to achieve statistical certitude in an experimental setting, experimenters are forced to limit extraneous causes of error, such as differences in playback equipment for the judges. These differences could muddy the listening capacity of musical professionals, and skew the results of this study.

Factors other than bit rate, sampling rate, and the amount of noise in a recording also affect the quality of the recording. The hall in which the recording takes place, ambient noise in the hall, microphone placement, audience noise, and many other factors all contribute to a successful recording. Although the interplay of these factors is out of the scope of this particular project, the study still proves that musicians can hear quality differences. With the extreme level of competition in audition situations, one would surmise that a performer would want every advantage possible, and a high-fidelity CD recording provides such an advantage.

References

Ando, Y. (1998). *Architectural Acoustics: Blending Sound Sources, Sound Fields, and Listeners*. New York: Springer-Verlag.

Bareham, J. R. (1996). Measurement of spatial characteristics of sound reproduced in listening spaces. *Audio Engineering Society Preprint, 101st Convention*, preprint no. 4381.

Bech, S. (1987). Planning of listening tests – choice of rating scale and test procedure, in Bech, S. and Pedersen O. J., eds. *Proceedings of a Symposium on Perception of Reproduced Sound*. Denmark: Stougaard Jensen. 61-70.

Berg, J. & Rumsey F. (2000). Correlation between emotive, descriptive and naturalness attributes in subjective data relating to spatial sound reproduction. *Audio Engineering Society Preprint,* 109th Convention, preprint no. 5206.

Blauert, J. & Lindemann, W. (1986). Auditory spaciousness: some further psychoacoustic analyses. *Journal of the Acoustical Society of America, vol. 80*, (2). 533-542.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, USA: MIT Press.

Gabrielsson, A., Hagreman, B., Bech-Kristensen, T. & Lundberg, G. (1990). Perceived sound quality of reproductions with different frequency responses and sound level, *Journal of the Acoustical Society of America, 88*, 1359-1366.

Grewin, C. (1995). Can objective measures replace subjective assessments? *Audio Engineering Society Preprints,* 99th Convention, preprint no. 4067.

Griesinger, D. (1997). The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces. *Acoustica, 83* (4). 721-731.

Griesinger, D. (2001). The psychoacoustics of listening area, depth, and envelopment in surround recordings, and their relationship to microphone technique. *Proceedings of the 16th*

International Audio Engineering Society Conference, Bavaria, Germany, 182- 200.

Guski, R. (1997). Psychological methods for evaluating sound quality and assessing acoustic information. *Acta Acustica, 83*. 765-774.

Hansen, V. & Munch, G. (1991). Making recordings for simulation tests in the Archimedes project. *Journal of the Audio Engineering Society, 39* (10). 768-774.

Harris, R. (2002). *Audiophilia*, November 2002.

International Telecommunications Union, (1997). Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. International Telecommunications Union - Radiocommunications, Recommendation ITU-R BS 1116.

Kirk, R. E. (1956). Learning, a major factor influencing preferences for high-fidelity reproducing systems. *Journal of the Acoustical Society of America, vol. 28* (6). 1113-1116.

Kosslyn, S. M. (1981). The medium and the message in mental imagery: a theory, *Psychological Review, 88* (1), 46-66.

Legge, A. (1990). *The Art of Auditioning*. London: Rhinegold.

Lipshitz, S. (1986). Stereo microphone techniques: are the purists wrong? *Journal of the Audio Engineering Society, 34*, (9). 716-744.

Mason, R. and Rumsey, F. (2000). An assessment of the spatial performance of virtual home theatre algorithms by subjective and objective methods. Audio Engineering Society, 108th AES Convention, preprint 5137.

Mapp, P. (1997). "Effects of Equalization on Sound System Intelligibility and Perceived Performance." 103rd AES Convention. New York.

McKinnie, D, (1996). *Objective Selection of Critical Material for Subjective Testing of Low Bit-rate AudioCoding Systems*. Master's Thesis, McGill University, Montreal.

McKinnie, D, (1991). Recording Techniques and the Perception of Environment, *Audio Engineering Society Preprint, 91st AES Convention*. Preprint No. 3110.

Meares, D. J. (1993). Perceptual Attributes of Multichannel Sound. *Proceedings of AES 12th International Conference 'The Perception of Reproduced Sound',* 171-179.

Meilgaard, M., Civille, G. V. and Carr, B. T. (1991). *Sensory Evaluation Techniques, 2nd edition,* Boca Raton, FL: CRC Press.

Møller, H., Sørensen, M. F., Jensen, C. B. and Hammershøi, D. (1996). Binaural technique: Do we need individual recordings? *Journal of the Audio Engineering Society, 44*, 451-469.

Moore, B. C. J. (1997). *An Introduction to the Psychology of Hearing, 4th edition.* London: Academic Press,.

Newell, P. R. & Holland, K. R. (1997). "A Proposal for a More Perceptually Uniform Control Room for Stereophonic Music Recording Studios." 103rd AES Convention. New York.

Ohashi, T., Nishina, E., Kawai, N., Fuwamoto, Y., & Imai, H., (1991). High Frequency Sound Above the Audible Range Affects Brain Electrical Activity and Sound Perception, AES 91st Convention, New York, preprint 3207.

Ohashi, T., Nishina, E., Fuwamoto, Y., & Kawai, N., (1993). On the Mechanism of Hypersonic Effect. *Proceedings Int'l Computer Music Conference*, Tokyo, 432–434.

Precoda, K. Meng, K. (1997). "Subjective Audio Testing Methodology and Human Performance Factors" 103rd AES Convention. New York.

Rumsey, F. (1998). Subjective assessment of the spatial attributes of reproduced sound. *Proceedings of the 15th International Audio Engineering Society Conference,* Copenhagen, Denmark, 122-135.

Rumsey, F. (1999). Controlled subjective assessments of two-to-five-channel surround sound processing algorithms. *Journal of the Audio Engineering Society, 47* (7/8). 563-582.

Sauer, M. (2000). *Stereophile, 1*, 57.

Schroeder, M. R. (1993). Listening with Two Ears. *Music Perception, 10* (3), 255–280.

Stuart, J.R. (1994). "Perceptual issues in multichannel environments" 97th AES Convention, San Francisco.

Stuart, J.R. (1991). Psychoacoustic models for evaluating errors in audio systems. *PIA, 13* (7), 11–33.

Toole, F. (1985). Subjective measurements of loudspeaker sound quality and listener performance, *Journal of the Audio Engineering Society 33*, (1/2). 2-32.

Terhardt, E., (1990). Music perception and sensory information acquisition: relationships and low-level analogies. *Music Perception, 8* (3), 217-239.

Umemoto, T. (1990). The Psychological Structure of Music. *Perception, 8* (2), 115–128.

Zielinski S. K., Rumsey F., & Bech S. (2002). Subjective audio quality trade-offs in consumer multichannel audio-visual delivery systems. Part I: Effects of high frequency limitation. AES 112th Convention, Paper 5562.