LISTENING THROUGH DIFFERENT EARS IN THE SYDNEY OPERA HOUSE

Angela Qian Li, Craig Jin, André van Schaik

The University of Sydney
School of Electrical and Information Engineering
Sydney, NSW Australia 2006

ABSTRACT

We present a psychoacoustic experiment that explores the ability of various listeners to discriminate between the virtual auditory space (VAS) stimuli generated using different binaural impulse response functions recorded in the Sydney Opera House. The binaural head-related impulse response (HRIR) functions were recorded for a group of subjects sitting in the same seat, P34, using a log sine sweep sound source located at the centre of the stage. The VAS stimuli generated using these HRIRs consist mostly of a variety of musical excerpts, speech, and white noise. Experimental results using an ABX test procedure show that out of a total of 1350 trials, 10 subjects responded correctly in 1230 of the test trials, indicating a discrimination performance greater than 90%. We also present data indicating the types of perceptual cues that aid in binaural sound discrimination process.

1. INTRODUCTION

It is generally accepted that individualized binaural recordings of music in a concert hall differ from listener to listener and that these differences are primarily due to the differences in the acoustic filtering of the outer ear. In other words, the binaural recordings made using one microphone in each ear for two listeners sitting in the same seat within the same concert hall listening to the same music would be different. We ask the question: how well can listeners discriminate these differences? Presumably, these differences are discernable and are the primary reason there are so few binaural recordings available of music in concert halls. What sounds correct to one listener may appear colored and distorted to another listener. Despite this fact, the perceptual quality of concert halls is frequently assessed using non-individualized binaural recordings [1].

Previous research in this area seems to concentrate primarily on subjective evaluations of the spatial quality of sounds, particularly for sound reproductions systems [2][3][4]. Particular emphasis is given in the published literature to the difficulty in characterizing spatial audio with typical approaches using descriptive analysis, semantic methods [5], and even a hierarchy of spatial attributes [3]. The work presented here takes a different approach in that we are not assessing the quality of a sound reproduction system per se, but examining the ability to discriminate binaural sounds corresponding to different listeners' ears. We also explore the auditory cues that enable subjects to discriminate these sounds.

2. METHOD

2.1. Binaural HRIRs

Binaural head-related impulse response (HRIR) measurements were conducted in the Concert Hall of the Sydney Opera House using a Soundsphere loudspeaker and a customized sub-frequency loudspeaker at six positions on the stage platform. Measurements were made in seat P34 for seven subjects and one acoustic mannequin which was a customized model (ears, head, and torso) of one of the authors of the paper [6]. The HRIR measurements were recorded on an Alesis HD24 Hard Disk Recorder using seven log sine sweeps from 20 Hz to 22 kHz with a duration of approximately five seconds. Subjects were instructed to directly face the center stage and sit squarely and aligned with the back of their chair.

After the binaural HRIRs were recorded, they were windowed to 2.6 seconds in length and their associated interaural time difference (ITD) value and interaural level difference (ILD) value were analyzed through listening tests and calculations using the MATLAB software. The primary issue is that a subject may not have been directly facing the centre stage leading to an 'artificial' ILD and ITD cue than can be used to discriminate between different binaural HRIRs. Thus, two subjects' HRIRs were removed from the data set, leaving a total of six binaural HRIRs. Additionally, as the pre-amplifier gain settings may not have been exactly matched for different subjects during the recordings, the HRIRs across different subjects were normalized to an average level using their power spectral density up to 1 kHz.

2.2. Sound stimuli

The six recorded binaural HRIRs were used to create nine VAS sound stimuli using nine sound excerpts taken from two archives of anechoic recordings of music and speech (see Table 1). A variety of sound stimuli were chosen ranging from speech, instrumental music, ensemble music, and orchestral music. The sound excerpts covered different frequency ranges and textural complexities and were limited to 10-15 seconds in length.

Anechoic Sound Source	Abbreviated Song Name	Track No
Music for Archimedes CD	Guitar	15, "Etude No.6 in E minor" by H.Villa Lobos
	Cello	20,"Theme" by Weber
	Xylophone	27," Sabre Dance" by Khachaturian
	Female voice	4, Female speech
Denon Professional Test CD 2	Flute	31
	Jazz ensemble	35
	Orchestra – Moz.	25, "Die Hochzeit Von Fiagro" by Mozart
	Orchestra – Sh.	26, "Symphony No.5" by Shostakovich
	White noise	Random generated broadband noise

Table 1: Nine mono anechoic excerpts are extracted from two CD archives, ranging from speech, tonal instruments to complex ensemble and orchestral pieces.

A total of 54 VAS sound stimuli were generated by convolving the binaural HRIR filters with the nine mono anechoic sound stimuli described above. The correctness of the processing steps described above were verified by comparing the simulated binaural VAS sound stimuli for the Orchestra-Moz. piece with a true binaural recording of the same piece that was recorded in the same session as when the binaural HRIRs were recorded. The 54 stimuli are made available at http://www.eelab.usyd.edu.au/andre/SOH/.

In order to achieve a consistent perceived loudness across the 54 VAS sound stimuli, a loudness model [7] was applied to the left and right ear sound signal for each sound stimulus. A single average gain adjustment factor that was required to match an 83 dB SPL pink noise was calculated and applied to the left and right ear sound signals. The sound stimuli were played to the listeners using the Sennheiser HD600 open headphones. For each listener participating in the experiment, a headphone transfer function was measured in an anechoic room and an inverse calibration filter function was generated using an adaptive least mean square algorithm (see Figure 1).

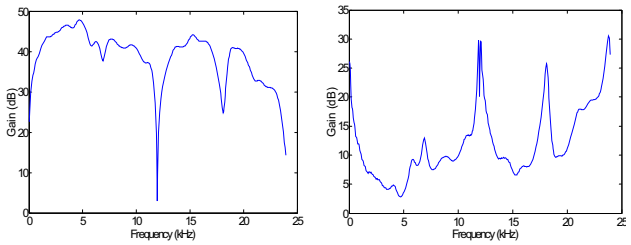

Figure 1. The plot shows recorded Sennheiser HD600 headphone impulse on the left and its calculated headphone inverse function on the right.

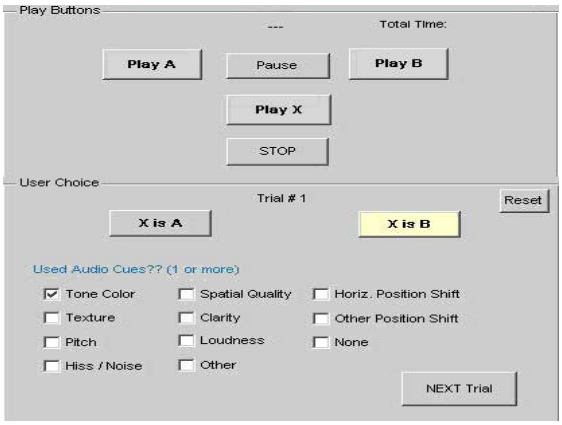

Figure 3. An ABX user interface design, where subjects have to select X is A or B and what audio cues they mostly used by ticking from 11 audio cues checkbox in each trial before proceeding to the next trial.

Ten subjects performed the ABX test – nine males and one female averaging 25.50 years of age, all with self-reported normal hearing. Four of the subjects had no previous experience in listening experiments.

The binaural impulse responses recorded in the Concert Hall of the Sydney Opera House for six individuals were used. These allow for 6(6-1)/2=15 different binaural pairing combinations. Nine different pieces of source material were used, resulting in 9 × 15 =135 trials per subject. These trials were divided into three sessions. Each session consisted of 45 trials (= 5 pairs × 9 songs) with random HRIR pairing and rotated through the nine songs. For training purpose, each subject is given up to 20 practice trials before the start of the experiment.

In each trial, the subjects were able to play the stimuli A, B, and X as often as they wanted before selecting whether X was equal to A or B. The subjects were also asked to check one or more of the audio cue boxes on the ABX user interface to indicate which cues best represented those used by the subject to distinguish between A and B on this trial. The audio cue options are given in Table 2 and were explained to each subject at the start of the first session.

Audio Cue	Description
Tone Color	Timbre color of sound, e.g. bright, warm, rich
Texture	Density of sound, instrumental layers
Spatial Quality	Envelopment of sound source, spatial cohesion, and source width
Hiss Noise	Audible background artifacts
Clarity	Cleanness, crispness of sound
Horizontal Position Shift	Sound image perceived at a different azimuth
Other Position Shift	Distance or elevation change
Pitch	Frequency change, sharp or flat
Loudness	Volume, intensity of sound
Other	None of the above 1-8 cues but able to detect differences
None	Assumes the subject has guessed this trial because he is unable to detect any differences

Table 2. The 11 psychoacoustic audio cue variables are described as above to all subjects participated in the experiment.

3. RESULTS

3.1. Overall subject performance

On the ABX tests, the ten subjects combined were able to correctly determine X on 1230 out of 1350 trials, which is equivalent to 91.11% percent of the trials. The individual score for each subject is given in Figure 4. Even though these results indicate that the subjects were quite sensitive to differences in the reverberant binaural impulse responses, these differences were quite subtle. Subjects reported that it was quite a difficult task, especially after the first session. A number of subjects needed more than 1.5 hours to complete a block of 45 trials. By the third session all subjects found that they had become more familiar with the music and were more aware of what to listen for. Subjects performed generally better and took less time to do the final two sessions than the first session, but many still needed more than one hour for the final block of 45 trials.

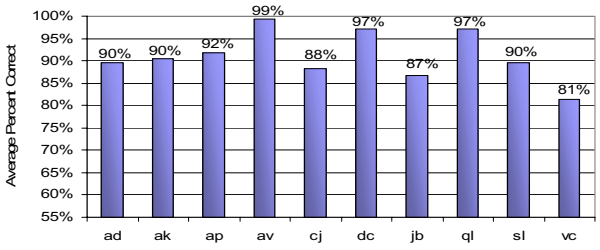

Figure 4. This graph shows the overall test score for each subject in 135 trials.

3.2. Breakdown of audio cues

Subjects also reported which cues from Table 2 they used for each ABX decision. Of the 11 psychoacoustic variables, the subjects mostly relied on "Tone Color" (21%), "Spatial Quality" (20%) and "Other Position Shift" (15%) as shown in Figure 5. This is reasonable as the same source material was used for all AB pairs, loudness was equalized for horizontal position shifts, level differences were mostly absent and noise was removed as much as possible. Since different HRIRs provide different spectral cues, variations in the impression of different spatial locations, quality and coloration of the source material can be expected.

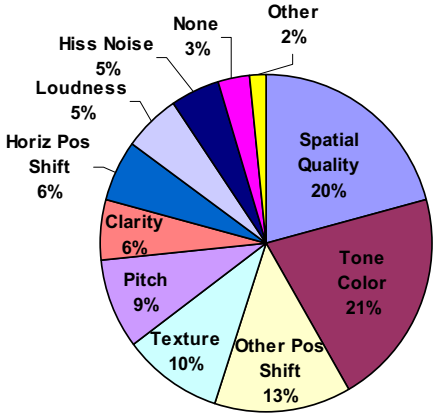

Figure 5. The fraction of each audio cue used by subjects in the experiment is shown in the pie chart as percentages.

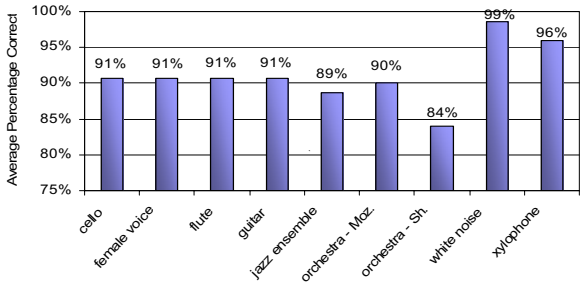

Figure 6. This graph shows the average percentage correct per song. Each song is played to 10 subjects 15 times with different binaural HRIR pairing.

3.3. Perception of complex music stimuli

Figure 6 shows the ABX scores for each of the nine pieces of source material. White noise (99%) was easiest excerpts to compare, with the subjects being able to correctly distinguish A and B on 99% of the trials. The xylophone piece was the next easiest for discrimination, with 96% of the trials correct. Both pieces only contain a single sound source (as opposed to orchestral music). Both were also broadband – white noise by definition and the xylophone due to the transient nature of its percussive sound. Other solo instrumental songs and the female voice were distinguished correctly on 91% of the trials. The spectra of these sources were less broadband in nature than the white noise or the xylophone. The jazz ensemble and the two orchestral excerpts scored lowest, with the Shostakovich piece being the most difficult. The Mozart piece, Marriage of Figaro, is played in unison and is simpler in instrumental textural layers

than the 20th century piece by Shostakovich's. This makes it much harder to concentrate on a single auditory element in the Shostakovich piece. One would expect the same difficulty with the jazz ensemble however, the high-hat drum beat and a flute solo provided strong single instrumental focal points in the music.

The subjects' average scores are given for each of the 15 possible HRIR pairings in Figure 7. The different recorded binaural HRIRs are indicated by letters A, G, R, B, J and H. It can be seen that the HRIRs for A and G are most similar as the subjects most easily confused these two. Next most similar are A and R. The HRIRs of H were most dissimilar to all of the other HRIRs, except maybe for those of R, and subjects could almost always discriminate correctly between the HRIRs of J and any other individual's HRIRs. These differences can be further depicted in the spectrogram of the convolved binaural HRIR with sound stimulus. The most dissimilar pair H and J shows significant difference in the higher frequencies in spectrogram in Figure 8.

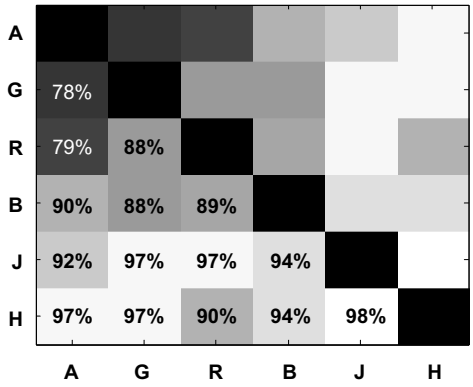

Figure 7. This grayscale grid map shows 15 possible different HRIR pairing combinations. Each grid represents the average percentage of correctly scored trials for that pair. The increase of percentage corresponds to lighter grid shades for dissimilar HRIR pairs and similar HRIR pairs are shown in darker grid shades.

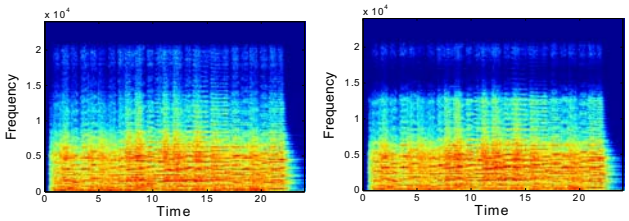

Figure 8. Spectrogram on stimuli: Orchestra – Sh. For the most dissimilar pair subject H (left) and J (right)

4. DISCUSSION

A recent subjective preference study [9] showed significant increase in spatial presence rating for individualized binaural stimuli over stimuli processed with generic HRIRs. Other studies have shown that applying non-individualized HRIRs results in three kinds of localization errors: front-back confusions, elevation error and inside-the-head error. A front-back confusion results when the listener perceives the sound to be in the front when it should be in back and vice-versa. An elevation error refers to a misperceived elevation angle, for example an overhead sound source may be perceived to be behind the listener and inside-the-head error when the sound source does not sound externalized from the head [10][11]. The main findings in the present study support the need for individualized binaural recordings or individualized HRIRs in constructing binaural VAS sounds. Differences between a listener's own ears and those of others lead to changes in the sound that are, while subtle, often perceptible.

5. REFERENCES

[1] P. Martignon, A. Azzali, D. Cabrera, A. Capra and A. Farina, "Reproduction of auditorium spatial impression with binaural and stereo phonic sound systems", Proceedings of the AES 118th Convention, Barcelona, 2005.

[2] A.H. Marshall and M. Barron, "Spatial Responsivenss in Concert Halls and the Origins of Spatial Impression", Applied acoustics 62, 2001.

[3] F. Rumsey, *Spatial Audio,* Focal Press, UK, pp 41-45. 2001.

[4] D. Grisinger, "General Overview of Spatial Impression, Envelopment, Localization, and Externalization", Proceeding of the AES 15th International Conference, Denmark, 1998.

[5] S. Bech, "Methods for Subjective Evaluation of Spatial characteristics of sound", Proceeding of the AES 16th International Conference, Finland, 1999.

[6] C. Jin, Spectral Analysis and Resolving Spatial Ambiguities in Human Sound Localization, PHD thesis, The University of Sydney, 2001.

[7] D. Robinson, A Proposed Standard for Equal Loudness Filter, http://www.replaygain.org/,

[8] N.A. Macmillan and Creelman C.D, *Detection Theory: A User's Guide*, Cambridge University Press, New York, 1996.

[9] A Valjamae, P. Larsson, D. Vastfjall and M. Kleiner, "Auditory Presence, Individualized Head-Related Transfer Functions, and Illusory Ego-Motion in Virtual Environments", Proc. Of 7th Annual Workshop Presence, Valencia, Spain, 2004.

[10] D.R. Begault, E.M. Wenzel, A.S. Lee and MR. Anderson, "Direct comparison of the impact of Head Tracking, Reverberation, and Individualized HRTFs on the Spatial Perception of a virtual Speech Source", AES 108th Convention, Paris, France, 2000.

[11] S. Carlile, "Virtual auditory space: Generation and applications", Austin, 1996.