VERTICAL IMAGING CAPABILITY OF SURROUND SOUND SYSTEMS THROUGH BINAURAL TECHNIQUES

John Klepko

Faculty of Music
McGill University
Montreal, Quebec
H3A 1E3

INTRODUCTION

Recent advances in audio technology have allowed storage and playback of multichannel recordings for music and video applications. With these advances comes a need to research new methods of encoding the soundfield that can take advantage of the enhanced spatial reproduction ability surround-sound has over conventional 2-channel stereophonic systems. The most prevalent multichannel systems in use involve 5 channels of information stored discretely and replayed over 5 separate loudspeakers. The typical loudspeaker layout is shown in Fig. 1.

Most of the recent research involving the imaging capacity of these systems is concerned with the horizontal plane surrounding the listener. Achieving controllable placement of sounds in the 360° surrounding the listener with only 5 channels is a formidable task in itself, but in order to more completely reproduce or simulate a real acoustical space, the vertical dimension should also be represented.

This paper proposes the use of a 5-point microphone array to encode an acoustical event with both horizontal and vertical imaging capabilities. Previous investigations [1] with this system have reported reasonably accurate horizontal imaging so the present study will focus on the vertical imaging.

Complete and accurate vertical imaging is not expected from this system since this is not even possible in natural hearing. Unlike the research in vertical localization, this investigation is based on creating an "illusion" of a sound source position. This is due to the fact that the *intended* vertical positions do not coincide with the *actual* sound sources as there are no loudspeakers above the listener during playback of the recording. As well, it is realized that the difficulty and errors which occur naturally in vertical localization can obscure the test results of this system.

The main objective of this experiment is to ascertain whether this system is able to resolve recorded sound events at these difficult overhead positions. A sub-objective is to discover whether certain sound stimulus types are more often correctly localized overhead than others.

METHOD

The recording apparatus consists of 5 microphones all from the same manufacturer. These were chosen because they use the same transducer element with equal sensitivity (25 mV/p) and noise specifications (10-12 dBA). This way, all microphones have virtually the same dynamic performance which is necessary for consistent level-matching across all 5 channels. Each microphone is assigned to a separate dedicated channel; the left and right are fed from identical super-cardioid pattern microphones, the centre channel from a cardioid microphone. The 2 surround channels are fed from a custom-made dummy-head microphone (with shoulders) incorporating 2 pressure-type omnidirectional microphones fitted into the ear molds. All 5 microphones are amplified by preamps of the same type built into the recording console. All gains are set to the same level (ie. no compensation) and subsequently played back and monitored at matched levels. All 5 channels are recorded to a digital tape-based 8-track at 16-bit word length and 48 kHz sampling rate. The playback setup (Fig. 1) for the experiment consists of 5 loudspeakers of the same type. They are simple 2-way direct radiator loudspeakers. All 5 loudspeakers are placed equidistant from the listening position at 2 metres.

In order to simplify the investigation at this initial stage, the test positions are limited to 5 positions across the median plane as shown in Fig. 2. The sound samples included 5-second samples of; male speech, 2 separate percussion instruments (shaker, sleigh bells) and 1/3-octave band filtered noise played through a small 2-way loudspeaker. The center frequencies were; 0.25, 0.5, 1, 2.5, 4, 6, 8, 10, 12.5 and 16 kHz. All 5-second samples were divided into 3 short, repeated phrases (or bursts). They were recorded with the 5-point microphone array from a distance of 1.5 metres. In an attempt to simulate a practical application of this system, the recording was performed on a large concert hall stage. These sound sample types and positions were randomly reorganized and re-recorded onto a test tape. The total number of sample points was 65 and the duration of the test was about 20 minutes. 5 subjects (with no reported hearing defects) were involved in the test. They were instructed to note down the perceived position (along the median plane) of the sound event. The subjects were seated but not restricted from moving.

FIGURE 1

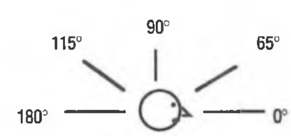

FIGURE 2

RESULTS

Figure 3 compares the 2 general sound types simplified into 3 positions (with above = 65°, 90° and 115°). Figure 4 compares the percentage of correct responses for the 5 different positions for all sound types. Figures 5 and 6 compare the percentage of correct responses for the natural sounds and narrow-band noise respectively.

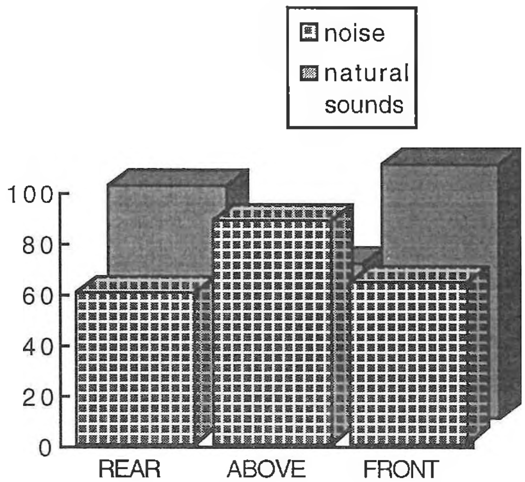

Figure 3

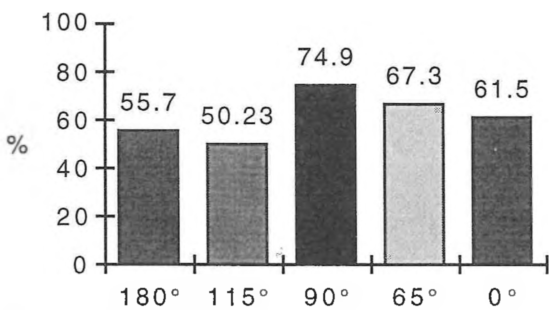

Figure 4

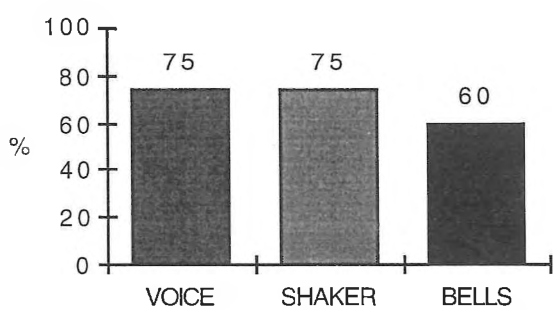

Figure 5

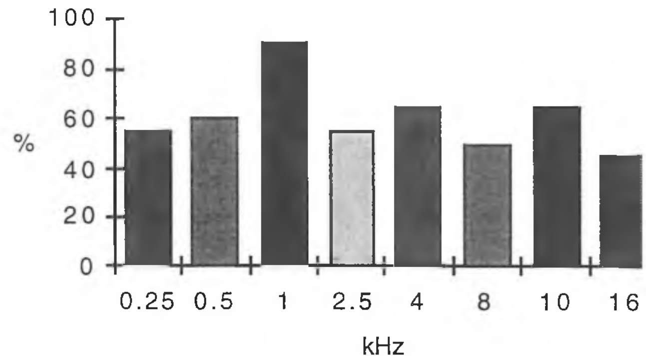

Figure 6

DISCUSSION

From these results it can be shown that it is possible to perceive overhead sounds despite there not being a sound source (loudspeaker) above the listener. The high value for the direct overhead position (Fig. 4) was due to the noise samples which were more often localized above. The natural sounds (i.e. voice, shaker, bells) did much worse for the overhead sounds, but scored excellent for the front and rear positions. The explanation for these results may be that the natural sounds are more familiar and are "expected" to be on ground level rather than above. The narrow-band noises are synthetic sounds which are unfamiliar. These intangible sounds carry no expectations possibly allowing the listener to believe the illusion of an overhead position. But it should also be noted that the natural sound types scored the highest overall percentages over the noise types which is due to the good localization for the front and back positions.

As expected, the low frequency noise bands were difficult to localize, especially overhead. But part of this can be explained by the natural hearing system's difficulty with such sounds. However, there was no trend towards better localization with increasing center frequency as reported by Gardner [2]. Perhaps the sound samples were too short. (Informal listening experiments with longer samples that were not randomized provided significantly better results).

CONCLUSION

A new method of encoding the soundfield in its total spatial dimension has been presented with particular attention to overhead imaging capability. Although not very precise, this investigation shows that the proposed system has the ability to represent overhead sounds to some controllable degree. More investigations of different sound types are planned.

REFERENCES

[1] KLEPKO, J., "5-Channel Microphone Array with Binaural Head for Multichannel Reproduction", AES Convention, New York, 1997, preprint # 4541.

[2] GARDNER, M., "Some Monaural and Binaural Facets of Median Plane Localization", JASA, vol. 54, no. 6, 1973.