Sound Localization and Virtual Auditory Space

Zhan Huan Zhou, *Member, IEEE*
Institute of Biomaterials and Biomedical Engineering
Edward S. Rogers Department of Electrical and Computer Engineering
University of Toronto, CANADA

***Abstract*—Many factors affect sound localization. Cues such as time and level differences between the ears as well as spectral information are thought to be important in localization. It is proposed that sound waves in the eardrum should produce the same effect on a listener regardless if the sound originated from free space or headphone delivered stimulus. A filter for headphone delivered stimulus to create a virtual auditory space was implemented and tested. Listeners could localize sound from the filtered headphone stimulus but it was not perfect. Other information arising from head movements, monaural cues and learning may enhance localization. Verification of the duplex theory is performed with a virtual auditory space.**

***Index Terms*—virtual auditory space, sound localization, virtual reality, surround sound**

I. INTRODUCTION

From home theatre systems to virtual reality, sound localization is essential for providing a rich auditory environment. Recent advancements in digital signal processing have allowed systems to be developed that can localize sound with only two free-field speakers. Localization can also be achieved using headphones by creating a "virtual auditory space." A virtual auditory space over headphones can be used in conjunction with virtual reality systems and potentially for hearing aids. This paper focuses on the development of a virtual auditory space with headphone delivered stimulus.

II. DUPLEX THEORY

The "duplex theory" was proposed by Lord Rayleigh at the turn of the century. It was thought that the auditory system localized sound in a similar fashion to the visual system. Two eyes localize an object in space so it was proposed that two ears should also localize sound. We have two ears separated by a relatively large head. In general, the onset time of a sound will be different for each ear. This is referred to as the interaural time difference (ITD). For continuous sounds, this is represented as an interaural phase difference (IPD). Pyschophyical experiments have shown that these localization cues are effective only in the range below 1.5kHz [1]. Another mechanism known as the interaural level difference (ILD) may be used. Since the head is a relatively dense medium, it will tend to cast an acoustical shadow on the ear contralateral to the sound source. The attenuation has been measured to be just over 40dB [1] for frequencies above 3kHz. However, for frequencies below this, the attenuation is almost zero and consequently, so is the ILD. Clearly, the duplex theory is incomplete. First, there is a frequency gap for which neither ITDs nor ILDs can be used as cues for localization. Secondly, an interaural time difference can be mapped to multiple regions in space, causing a "cone of confusion." A sound originating on the medial plane should have zero difference while a sound on the interaural axis will produce a constant time difference, regardless of distance. Third, it has been shown sounds can be localized with only monaural cues [2]. This suggests that physical cues other than interaural time and level differences are used in sound localization. It has recently been proposed that pinna filtering provides spectral cues that can be used to localize sound. Differences in filtering at each ear may provide cues to resolve spatial origin of sounds that lie on the cone of confusion. Recent experiments with headphone delivered stimulus has helped to uncover some of these cues and the information has been used to create a virtual auditory space.

III. HEAD-RELATED TRANSFER FUNCTION

A. Theoretical Derivation

A basic assumption in the creation of a virtual auditory space is that if the acoustical waveforms present at a listener's eardrums are the same under headphones as in free field, then the listener's experience should also be the same [3]. Since it is proposed that pinna filtering provides spectral cues for sound localization, linear systems analysis in the frequency domain with Fourier transforms is typically used to generate a "Free-Field-to-Eardrum Transfer Function" (FETF), also known as a "Head-Related Transfer Function" (HRTF). The terms are used interchangeable in the literature, however, HRTF is the common term used in the context of psychophysics and thus will be used for the remainder of this paper.

Typically, sounds generated from headphones appear to originate from within the head. In the virtual auditory space, the headphones should be able to "externalize" the sound. Using the HRTF, sounds can be spatially positioned using the technique described below (adapted from [3]).

Let $x_1(t)$ represent and electrical signal driving a loudspeaker and $y_1(t)$ represent the signal received by a microphone inside the listener's eardrum. Similarly, let $x_2(t)$ represent the electrical signal driving a headphone and $y_2(t)$ represent the microphone response to the signal. The goal of the virtual auditory space is to choose $x_2(t)$ such that $y_2(t) = y_1(t)$. Applying the Fourier transform to these signals, we come up with the following two equations:

$$Y_1=X_1LFM, \text{ and} \tag{1}$$

$$Y_2=X_2HM, \tag{2}$$

where L is the transfer function of the loudspeaker in the free field, F is the HRTF, M is the microphone transfer function, and H is the headphone-to-eardrum transfer function. Setting $Y_1 = Y_2$, and solving for X_2 yields

$$X_2 = X_1LF/H. \tag{3}$$

By observation, the desired transfer function is

$$T=LF/H. \tag{4}$$

Therefore, theoretically, if $x_1(t)$ is passed through this filter and the resulting $x_2(t)$ is played on the headphones, it should produce the same signal at the eardrum. Since the filter applies only to a single ear, another one must be derived for the other ear. This process is repeated for many places in the virtual environment to create an array of head-related transfer functions for each position to be recreated.

B. Practical Derivation

To obtain the head-related transfer function, a microphone is placed in the listener's eardrum such that it captures direction dependent effects and avoid the effects of standing-wave nulls at high frequencies. In the experiment by Wightman & Kistler [3], they generated head-related transfer functions from 144 distinct sound sources on a spherical surface. The stimulus was a broadband signal having spectral components from 200Hz to 14kHz with a duration of 20.48ms. The subject was instructed to hold their head still. A similar stimulus was output to the headphones. The frequency domain representation of the signals was obtained using a fast-Fourier transform. Since the experiment used the same stimulus ($X_1 = X_2$), dividing equation (1) by (2) results in the desired transfer function, $Y_1/Y_2 = LF/M = T$.

C. Quantitative Verification

Verification of the model is first done by measuring the "acoustical correctness" of the headphone delivered signal, that is, the sound wave in the ear canal generated from the headphones is the same as that produced from a free field loudspeaker. This was done by measuring an HRTF from a free field loudspeaker. A special FIR filter was constructed from this HRTF with a transfer function F/H. Using the impulse response of this filter as the input to the headphones, all the terms should cancel, except for the HRTF thus the Fourier transform of the recorded signal should exactly equal the HRTF obtained from the free field loudspeaker. This experiment was performed for four different source positions. Remarkably, the measured HRTF from the headphones varied only 1-2dB and 10^{o} in phase from the HRTF from that generated by the free field loudspeaker [3]. The most important result is that the error was independent of source position.

D. Psychophysical Verification

Although the acoustical correctness criteria described above is a quantitative measure of the effectiveness of the procedure, only psychophysical results can verify that the headphone stimulus produces the effect of a localized sound.

In an anechoic chamber, listeners were blindfolded and asked to identify the apparent location of a noise burst generated from an array of free-field loudspeakers while holding their head still as a control. Next, the test was repeated, except that the noise burst was delivered over headphones. The noise burst was filtered using the listener's personalized HRTF. The listener was again asked to identify the apparent location of the sound. Figure 1 shows the result of the experiment for two subjects. A straight line with a slope of one represents the ideal response. It is evident that in both the free field and headphone case, the listeners can judge azimuth reasonably well, while there is a slight confusion in determining elevation. In general, it appears that precision is greatest on the side (azimuth of $\pm 90^{o}$), slightly poorer in front, and poorest at high elevations in the rear [4]. The data also indicates that the correlation between target and response position is lower for the headphone condition and is more pronounced in source elevation. From these results, it is clear that that headphone does not capture all aspects of free-field hearing.

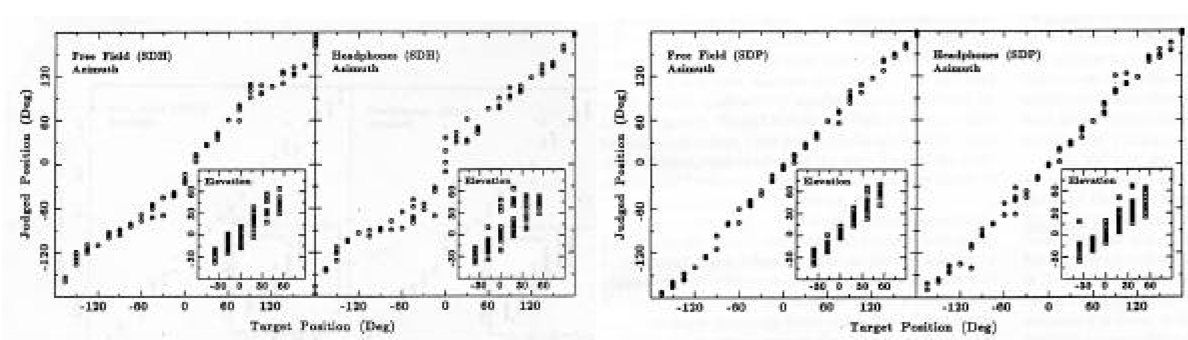

Figure 1 Comparison of localization in free-field and virtual auditory space for two subjects

E. Discussion

The psychophysical experiment validated that headphone delivered stimulus can provide the effect of a free field source. However, it does not capture the entire essence of the free field, especially in reproducing elevation effects. Although subjects were instructed to keep their head still in the free-field condition, subtle movements may have contributed cues to source localization. Head movements have no effect with headphone stimulus and thus may account for poor resolution of elevation in this case.

The experiment was conducted in a very specialized environment that is not experienced in the world at large. Stimulus was delivered in an anechoic room with a single sound source of known spectral content. In reality, there are many sound sources of unknown spectral content. Also, head movements are thought to provide localization cues while movements were restricted in this experiment. Perhaps the most important difference from the real world is the lack of auxiliary sensory data, especially visual feedback. Obviously, headphones cannot deliver all this information and thus are limited strictly to auditory cues.

One severe limitation is that the HRTF for headphone delivered stimulus must be individualized for the listener. It is not known whether a generic HRTF can be developed for commercial use. It is viable for applications not requiring many users, such as aviation training.

IV. OTHER LOCALIZATION CUES

A. Monaural Cues

A popular method of verifying the validity of the duplex theory is the study of monaural cues [5]. Typically, experiments were conducted by occluding one ear with putty. However, the effectiveness of such methods is questionable [2]. It has been shown that some sound energy is conducted through bones to the occluded ear. Although the attenuation is about 45 dB [6], the occluded ear will still detect residual sound energy, thus true monauralization is not achieved.

Using headphone delivered stimuli, Wightman & Kistler achieve almost true monauralization. Using head-related transfer functions, Wightman & Kistler delivered sounds from a virtual auditory space via headphones with one side disconnected. Their results were surprising. They found that localization was almost non-existent in the monaural virtual source condition [2]. However, although the sounds could not be localized, listeners reported that the sounds appeared to have an external origin.

Further experiments with free-field speakers indicated that monaural spectral cues are important for front-back and up-down perception. In this case, stimuli of a fixed spectrum was delivered to a listener. However, when the spectrum was randomized, the ability to resolve front-back and up-down disappeared. This suggests that *a priori* knowledge of spectral cues is important for deconvolution in monaural conditions since localization was still present in binaural conditions.

These experiments confirm that we process monaural cues, however, the spectrum of the sound must be known *a priori* for correct localization.

B. Head and Source Movement

A very natural response to localize a sound is a slight head movement. It is proposed that the slight shifts in auditory cues allow the brain to accurately localize a sound source. In sound localization experiments, the most common confusion experienced by listeners is front-back ambiguity. Slight head movements by the listener almost always resolves such ambiguities [7]. Wightman & Kistler tested this in a virtual auditory space. Using head tracking hardware, headphone stimulus was filtered to simulate a stationary sound source when a user moved his head. They found that even with headphone stimulus, front-back ambiguities disappeared with head movements.

In a secondary experiment, listeners were asked to keep their head still as the experimenters moved the apparent location of the source in the virtual space. To their surprise, they found that there were still a large number of front-back confusions. A third experiment was conducted similar to the previous one, but this time the listener was in control over movement of the source. In this case, front-back ambiguities once again disappeared.

Once again, this experiment demonstrates the importance of *a priori* knowledge to proper sound localization.

C. Effects of Learning

The importance of *a priori* knowledge to sound localization may reside in our ability to learn. In the head and source movement experiment, some listeners were asked why they didn't move their head. Some responded, "I don't need to move my head, and it doesn't help when I do." This suggests that certain individuals learn to estimate the spectral cues they are expected to encounter. In the case of monaural hearing, listeners who are deaf in one ear can localize sound much more accurately than "monauralized" binaural listeners [2]. This is probably due to the fact that the deaf listeners have much more experience extracting information from monaural cues. Wightman & Kistler also attempted to use someone else's HRTF on a listener. They found that localization was poor, but listeners could still externalize sounds [3].

V. CONCLUSIONS

Clearly, localization is not isolated to simply the sounds heard. Many more effects contribute to localization than that proposed by the duplex theory. Although Wightman & Kistler have shown that a virtual auditory space can be generated through headphone delivered stimulus, they are still lacking some key features. The ability to accurately reproduce elevation localization may be a problem for aircraft simulations. Other cues such as head movements and learning may also help in sound localization. For commercial applications where localization does not need such accuracy, an average HRTF can be created to externalize sounds.

REFERENCES

[1] S. Carlile, *Virtual Auditory Space and Applications,* Austin, TX, Springer, 1996.

[2] F.L. Wightman and D.J. Kistler, "Monaural sound localization revisited," J. Acoust. Soc. Am., vol. 101, pp. 1050-1063, February 1997.

[3] F.L. Wightman and D.J. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis," J. Acoust. Soc. Am., vol 85, pp. 858-867, February 1989.

[4] F.L. Wightman and D.J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," J. Acoust. Soc. Am., vol 85, pp. 868-878, February 1989.

[5] J.W. Strutt, "On our perception of sound direction," Philos. Mag. vol. 13, pp. 214-232.

[6] J.D. Hood, "Bone conduction: A review of the present position with especial reference to the contributions of Dr. Georg. Von Bekesy," J. Acoust. Soc. Am., vol. 91, pp 1325-1332.

[7] F.L. Wightman and D.J. Kistler, "Resolution of front-back ambiguity in spatial hearing y listener and source movement," J. Acoust. Soc. Am., vol. 105, pp. 2841-2853, May 1999.