Efficient Listening with Two Ears: Dichotic Time Compression and Spatialization

Barry Arons
MIT Media Laboratory
20 Ames Street
Cambridge MA, 02139
Email: barons@media.mit.edu

Abstract

To increase the amount of information we can collect in a given amount of time, it is possible to employ signal processing techniques to speed up the rate at which recorded sounds are presented to the ears. Besides simply speeding up the playback, it is possible to auditorily display the signals in a way that allows us to process and interpret the signals more efficiently by exploiting the use of our two ears.

This chapter first reviews time compression techniques for increasing the amount of information that can be presented to a listener, with an emphasis on techniques that use two ears. The chapter then describes a new technique that integrates these dichotic time compression techniques into a spatial audio display system.

1 Introduction

Auditory information is collected through our ears at a fixed rate and processed in our brain. To increase the amount of information we can collect in a given period of time, it is possible to employ signal processing techniques to speed up the rate at which recorded sounds are presented to a listener. These "time compression" or "time scale modification" algorithms have primarily been used on speech recordings. Besides simply speeding up the playback, it is possible to auditorily display the signals in a way that allows us to process and interpret the signals more efficiently by exploiting the use of our two ears. These "dichotic" time compression techniques present different portions of the audio signal to each ear, increasing intelligibility.[1]

Current spatial audio display systems attempt to take advantage of the fact that human listeners have two ears by creating virtual sound sources that are synthesized over headphones. However, one of the fundamental design premises of a spatial audio system conflicts with the presentation needs of a dichotic time compression algorithm. This prevents the use of the dichotic time compression technique with a conventional spatial audio system.

This chapter first reviews time compression algorithms for increasing the amount of information that can be presented to a listener, with an emphasis on methods that use two ears. The chapter then describes a new technique that integrates these dichotic time compression techniques into a spatial audio display system to further increase the bandwidth of the listener.

[1] *Dichotic* refers to two different signals that are presented to the ears over headphones.

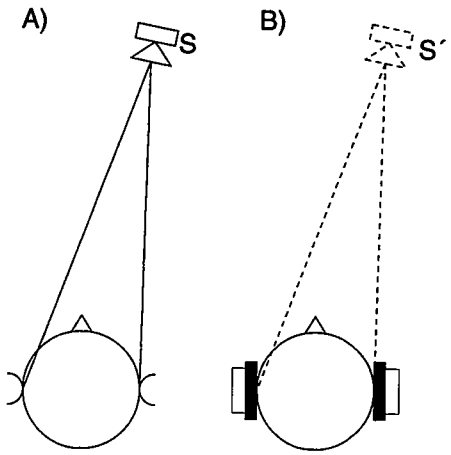

Figure 4: (A) Top view of a listener's head and sound source S (loudspeaker) located in space. (B) Virtual sound source S′ created by a spatial audio system.

5 Presenting DTCS Spatially

In spatial audio display systems one or more channels of audio are presented to the ears based on the head related transfer function (HRTF) and the spatial location of the source relative to each ear [16, 17]. For example, in Figure 4(A), a real sound source S is filtered based on the reflective characteristics of the head, body, and ears (pinna) and the interaural time delay due to the path length difference to the ears to produce a virtual sound S′ when presented over headphones (Figure 4(B)).

It is useful to be able to present time compressed speech in a virtual acoustic display, such as in user interfaces that allow skimming or browsing of recorded audio material [8, 19], or systems that attempt to present multiple streams of recorded speech simultaneously [20, 21]. Presenting speech that has been time compressed using the basic sampling or SOLA techniques in a spatial audio display system is straightforward, as it can be treated like any other audio source. However to exploit the improved intelligibility of dichotically presented time compressed speech within a spatial audio system a novel approach must be taken to spatialize DTCS.

The goal of DTCS is to explicitly present different signals to each ear, while a spatial audio system simulates a source at some spatial position by carefully controlling interaural time and intensity differences as well as the monaural spectral cues in the signals reaching the two ears. These two goals and their associated acoustic cues are thus seemingly in conflict. For example, if both channels of a DTCS signal are placed at the same location in a spatial audio system (e.g., at S′ in Figure 4(B)), both ears will receive a portion of the signal from each channel. Unfortunately, this cross talk will degrade the DTCS signal. As Gerber notes, "if one listens to both signals with both ears, the intelligibility is poorer than if one listens to one signal with one ear and the other signal with the other ear" [14] [p. 459].

However, it is possible to create a virtual sound source where each ear only receives one channel of the DTCS signal. This can be achieved by placing two virtual sound sources at the same location, but only filtering each signal for one ear (Figure 5(A)). One system configuration for creating this type of auditory display is shown in Figure 5(B).

Moore says that "while the most reliable cues used in the localization of sounds depends upon a comparison of the signals reaching the two ears, there are also phenomena of auditory space

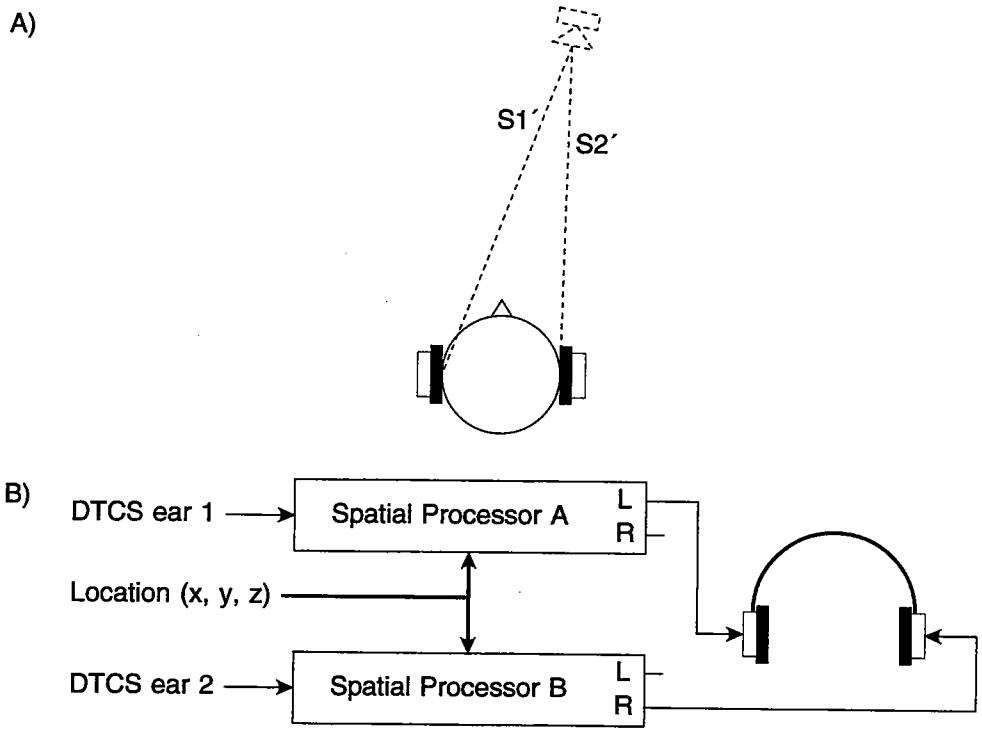

Figure 5: (A) Technique to present spatialized DTCS. S1′ and S2′ are two channels of DTCS originating from a single virtual sound location, but each is only presented to a single ear. (B) System configuration for spatializing DTCS using two Beachtron boards [22].

perception which result from monaural processing of the signals" [[9] p. 194]. Note however, that the spatialized DTCS technique described here is not strictly presenting two monaural channels, rather there are still a variety of rich interaural cues. The HRTF cues, including interaural intensity differences and monaural spectral cues, are all present. The only cue that is missing is interaural time difference, since the signals received by the ears do not originate from a single audio signal (however the two DTCS signals are still are highly correlated). Speech sounds in particular are very rich in familiar information. These common speech spectral cues make it easier for us to perceive these two channels as a single auditory stream [10].

This spatialized DTCS technique was informally found to produce an externalized virtual image. As with the DTCS technique, the speech sounds a bit choppy; however the speech was intelligible and comprehensible and could be localized about as well as a spatialized version of the original speech recording (time compressed, but not dichotic).

6 Issues

The work presented in this chapter has only scratched the surface of the spatialized DTCS technique. Further development of the underlying technique is needed as well as a formal evaluation to test the efficacy of this method of auditory display. Specific areas of research include: optimizing the time difference between the DTCS channels; exploring the perceived spatial and comprehension effects of permitting a small amount of crosstalk between channels (thus adding interaural time differences); and modifying the underlying spatial audio system architecture to allow DTCS to be presented spatially without requiring the use of two separate sound processing channels. The system also needs to be tested to perceptually evaluate if the sounds can be localized and externalized; if the maximum preferred time compression is degraded when the DTCS is spatialized; and if presenting spatialized DTCS enhances or hinders a listener's ability to listen to multiple audio streams.