Binaural Cue Selection for Source Localization: Motivation and Experiments

Christof Faller
Audiovisual Communications Laboratory, EPFL, 1015 Lausanne, Switzerland, Email: christof.faller@epfl.ch

Abstract

Recently, a binaural model for source localization in complex listening situations has been proposed (Faller and Merimaa, 2004). Concurrently active sources and reflections have the effect that the binaural cues at the ear entrances often do not anymore directly relate to the source directions. The values of the binaural cues are determined by the superposition of the various sources and reflections. The proposed model assumes that for localizing sources the binaural cues are only used at time instants when they determine a direction of a source correctly. These time instants are determined with the help of the interaural coherence which is continuously computed. A number of simulations are presented, including a scenario with concurrent speech and determination of an echo threshold.

Introduction

In most listening situations, the perceived directions of auditory events coincide with the directions of the corresponding physical sound sources. In everyday complex listening scenarios, sound from multiple sources, as well as reflections arrive concurrently from different directions at the ears of a listener. Thus, the auditory system does not only need a capability localize concurrently active sources, but it also needs to be able to ignore the effect of reflections on localization. Recently, we proposed a modeling mechanism explaining concurrent source localization and localization suppression of reflections [1]. In this paper, we are motivating this model, presenting simulation results, and new results on simulating echo thresholds.

Localization accuracy in the presence of concurrent sound has been investigated in a number of studies. A detailed review has been given in [2]. There has also been a number of recent studies on the effect of independent distracters on the localization of a target sound (see references in [1]). In order to understand localization of a source in the presence of reflections, the precedence effect [2, 3, 4] has to be considered. An important aspect of the precedence effect is localization dominance, which has also been often denoted the "law of the first wavefront". Depending on stimulus properties and individual listeners, localization dominance indicates that a leading stimulus dominates localization over a lagging stimulus for lead/lag delays between $2-50$ ms.

The Proposed Model

The auditory system features a number of physical, physiological, and psychological processing stages for accomplishing the task of source direction discrimination and ultimately the formation of the auditory spatial image. There is little doubt about the first stages of the auditory system, i.e. the physical and physiological functioning of the outer, middle, and inner ear are known and understood to a high degree. The first stages of the proposed model are based on models of the outer, middle, and inner ear. For the binaural processor, for the sake of generality, not a specific process is assumed, but it is assumed that the binaural processor yields to the upper stages of the auditory system information about the interaural time difference (ITD), interaural level difference (ILD), and interaural coherence (IC). The model is based on analysis of continuously computed ITD, ILD, and IC in critical bands.

Exponentially decaying estimation windows with a time constant of 10 ms are used for computing the $\mathrm{ITD}(n)$, $\mathrm{ILD}(n)$, and $\mathrm{IC}(n)$ in each critical band, where n is the time index. This time constant is in accordance with the temporal inhibition of the model of Lindemann [5].

Consider the simple case of a single source in free-field. Whenever there is sufficient signal power, the source direction determines the nearly constant ITD and ILD which appear between each left and right critical band signal with the same center frequency. The (average) ITDs and ILDs occurring in this scenario are denoted "free-field cues" in the following. The free-field cues of a source with an azimuthal angle ϕ are denoted ITD_ϕ and ILD_ϕ. It is assumed that this kind of a one source free-field scenario is the reference for the auditory system. That is, in order for the auditory system to perceive auditory events at the directions of the sources, it must obtain ITD and/or ILD cues similar to the free-field cues corresponding to each source that is being discriminated. The most straightforward way to achieve this is to select the ITD and ILD cues at time instants when they are similar to the free-field cues. In the following it is shown how this can be done with the help of the IC.

The time instants when $\mathrm{ITD}(n)$ and $\mathrm{ILD}(n)$ are similar to free-field cues, are selected as a function of $\mathrm{IC}(n)$, i.e. the $\mathrm{ITD}(n)$ and $\mathrm{ILD}(n)$ cues are only considered when $\mathrm{IC}(n)$ is larger then a certain threshold IC_0:

$$\mathrm{IC}(n) > \mathrm{IC}_0\,. \tag{1}$$

The same cue selection method is applicable for deriving the direction of a source while suppressing the directions of one or more reflections.

It is assumed that the auditory system adapts IC_0 for each specific listening situation, i.e., for each scenario with a constant number of active sources at specific locations in a constant acoustical environment. Since the listening situations do not usually change very quickly, it is assumed that IC_0 is adapted relatively slowly in time. In [1], it is also argued that such an adaptive process may be related to the buildup of the precedence effect.

Simulations

An example of the described cue selection process is shown in Fig. 1 for two speech signals arriving from $\pm 40°$ azimuth. $IC(n)$, $ILD(n)$, and $ITD(n)$ as a function of time are shown in the critical bands at 500 Hz and 2000 Hz. $ILD(n)$ and $ITD(n)$ are marked bold whenever $IC(n)$ is larger than the threshold IC_0. Note that the selected cues are always similar to the free field cues (marked as dashed lines in the figure).

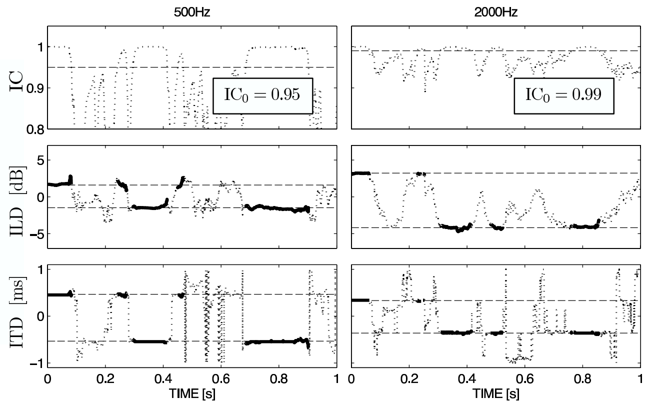

Figure 1: IC, ILD, and ITD as a function of time for two independent speech sources at $\pm 40°$ azimuth. Left column: 500 Hz, and right column: 2 kHz critical band. The cue selection thresholds (top row) and the free-field cues of the sources (middle and bottom rows) are indicated with dashed lines. Selected cues are marked with bold solid lines.

Other simulations, presented in [1], indicate that the cue selection can also model aspects related to the precedence effect. In the following, we are presenting some new results, related to modeling echo thresholds.

For the echo threshold simulations, the cue selection threshold IC_0 was computed as described in [1] (p. 3084), i.e. IC_0 was computed such that the variance of the selected ITD cues was smaller than 15 μs. To decide if there is an echo or not, we compared the ratio of the power of the selected cues for lead and lag to the corresponding ratio considering all cues (as a function of distance, cues are either assigned to lead or lag). If this ratio was larger than 30 dB it was assumed that no echo is perceived.

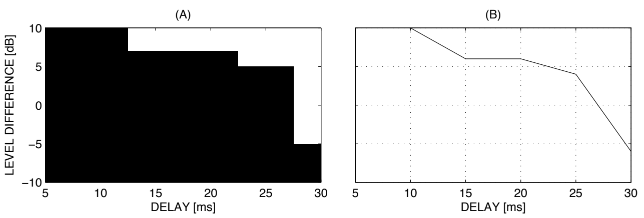

Figure 2: Computation of an echo threshold: Condition true (black) or false (white). (B) Corresponding echo threshold.

An echo threshold can be computed by examining the described condition for lead and lag signals with different level differences and delays. A result of such an analysis is illustrated in Fig. 2. Part (A) indicates where for which lead/lag pairs the echo is perceived and for which it is not perceived. Part (B) shows the corresponding echo threshold.

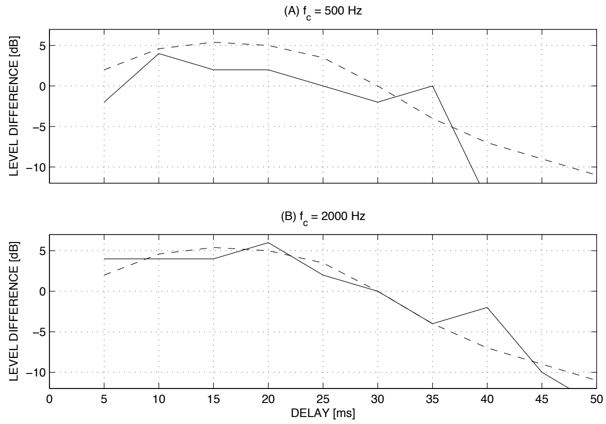

Figure 3: The computed echo threshold for critical bands at 500 Hz (A, solid) and 2000 Hz (B, solid) and the corresponding actual echo threshold [6] (dashed).

Figure 3 shows the echo threshold computed as described for a female speech signal for critical bands at 500 Hz (A) and 2000 Hz (B). As in [6] we used source angles $\phi_1 = 40°$ and $\phi_2 = -40°$. The computed echo threshold has a similar shape as the actual echo threshold given in [6].

References

[1] C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *J. Acoust. Soc. Am.*, vol. 116, no. 5, pp. 3075–3089, Nov. 2004.

[2] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press, Cambridge, Massachusetts, USA, revised edition, 1997.

[3] P. M. Zurek, "The precedence effect," in *Directional Hearing*, W. A. Yost and G. Gourevitch, Eds., pp. 85–105. Springere-Verlag, New York, 1987.

[4] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," *J. Acoust. Soc. Am.*, vol. 106, no. 4, pp. 1633–1654, Oct. 1999.

[5] W. Lindemann, "Extension of a binaural cross-correlation model by means of contralateral inhibition, I Simulation of lateralization of stationary signals," *J. Acoust. Soc. Am.*, vol. 80, no. 6, pp. 1608–1622, Dec. 1986.

[6] E. Meyer and G. R. Schodder, "Über den Einfluss von Schallrückwürfen auf Richtungslokalisation and Lautstärke bei Sprache," *Nachr. Akad. Wiss. Göttingen*, vol. 6, pp. 31–42, 1979, Math. Phys. Klasse IIa, (See Blauert 1997, p. 226).