Source localization in complex listening situations: Selection of binaural cues based on interaural coherence

Christof Faller[a)]
Mobile Terminals Division, Agere Systems, Allentown, Pennsylvania

Juha Merimaa[b)]
Institut für Kommunikationsakustik, Ruhr-Universität Bochum, Germany

(Received 27 January 2004; revised 19 July 2004; accepted 19 July 2004)

In everyday complex listening situations, sound emanating from several different sources arrives at the ears of a listener both directly from the sources and as reflections from arbitrary directions. For localization of the active sources, the auditory system needs to determine the direction of each source, while ignoring the reflections and superposition effects of concurrently arriving sound. A modeling mechanism with these desired properties is proposed. Interaural time difference (ITD) and interaural level difference (ILD) cues are only considered at time instants when only the direct sound of a single source has non-negligible energy in the critical band and, thus, when the evoked ITD and ILD represent the direction of that source. It is shown how to identify such time instants as a function of the interaural coherence (IC). The source directions suggested by the selected ITD and ILD cues are shown to imply the results of a number of published psychophysical studies related to source localization in the presence of distracters, as well as in precedence effect conditions. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1791872]

PACS numbers: 43.66.Qp, 43.66.Pn, 43.66.Ba [AK] Pages: 3075–3089

I. 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 from the surfaces of the physical surroundings, arrive concurrently from different directions at the ears of a listener. The auditory system does not only need to be able to independently localize the concurrently active sources, but it also needs to be able to suppress the effect of the reflections. In this paper, a modeling mechanism is proposed to explain both of these features. Before describing this modeling mechanism in more detail, related psychophysical localization experiments and psychoacoustic models are reviewed.

Localization accuracy in the presence of concurrent sounds from different directions has been investigated by several authors. A detailed review is given by Blauert (1997). The effect of independent distracters on the localization of a target sound has been recently studied by Good and Gilkey (1996), Good *et al.* (1997), Lorenzi *et al.* (1999), Hawley *et al.* (1999), Drullman and Bronkhorst (2000), Langendijk *et al.* (2001), Braasch and Hartung (2002), and Braasch (2002). The results of these studies generally imply that the localization of the target is either not affected or only slightly degraded by introducing one or two simultaneous distracters at the same overall level as the target. When the number of distracters is increased or the target-to-distracter ratio (T/D) is reduced, the localization performance begins to degrade. However, for most configurations of a target and a single distracter in the frontal horizontal plane, the accuracy stays very good down to a target level only a few dB above the threshold of detection (Good and Gilkey 1996, Good *et al.* 1997, Lorenzi *et al.* 1999). An exception to these results is the outcome of the experiment of Braasch (2002), where two incoherent noises with exactly the same envelope were most of the time not individually localizable.

In order to understand the localization of a source in the presence of reflections from different directions, the precedence effect needs to be considered. Extensive reviews have been given by Zurek (1987), Blauert (1997), and Litovsky *et al.* (1999). The operation of the precedence effect manifests itself in a number of perceptual phenomena: fusion of subsequent sound events into a single perceived entity, suppression of directional discrimination of the later events, as well as localization dominance by the first event. The directional perception of a pair of stimuli with an interstimulus delay shorter than 1 ms is called summing localization. The weight of the lagging stimulus reduces with increasing delay up to approximately 1 ms, and for delays greater than that the leading sound dominates the localization judgment, although the lag might never be completely ignored. Echo threshold refers to the delay where the fusion breaks apart. Depending on stimulus properties and individual listeners, thresholds between 2–50 ms have been reported in the literature (Litovsky *et al.*, 1999).

Localization accuracy within rooms has been studied by Hartmann (1983), Rakerd and Hartmann (1985, 1986), and Hartmann and Rakerd (1989) (see also a review by Hartmann, 1997). Overall, in these experiments the localization performance was slightly degraded by the presence of reflections. Interestingly, using slow-onset sinusoidal tones and a single reflecting surface, Rakerd and Hartmann (1985) found

[a)]Current address: Guetrain 1, CH-8274 Tägerwilen, Switzerland; Electronic mail: cfaller@agere.com
[b)]Also at Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology.

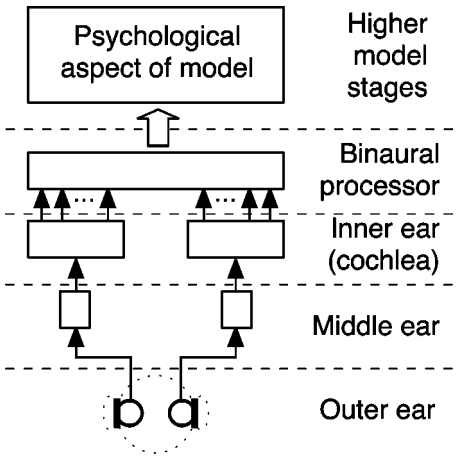

FIG. 1. A model of spatial hearing covering the physical, physiological, and psychological aspects of the auditory system.

that the precedence effect sometimes failed completely. In a follow-up study, the relative contribution of the direct sound and the steady state interaural cues to the localization judgment was found to depend on the onset rate of the tones (Rakerd and Hartmann, 1986). Nevertheless, absence of an attack transient did not prevent the correct localization of a broadband noise stimulus (Hartmann, 1983). Giguère and Abel (1993) reported similar findings for noise with the bandwidth reduced to one-third octave. Rise/decay time had little effect on localization performance except for the lowest center frequency (500 Hz), while increasing the reverberation time decreased the localization accuracy. Braasch *et al.* (2003) investigated the bandwidth dependence further, finding that the precedence effect started to fail when the bandwidth of noise centered at 500 Hz was reduced to 100 Hz.

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. The structure of a generic model for spatial hearing is illustrated in Fig. 1. 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. However, already the stage of the binaural processor is less well known. Different models have used different approaches to explain various aspects of binaural perception. The majority of proposed localization models are based on an analysis of *interaural time difference* (ITD) cues using a coincidence structure (Jeffress, 1948), or a cross-correlation implementation that can be seen as a special case of the coincidence structure. Evidence for cross-correlation-like neural processing has also been found in physiological studies (Yin and Chan, 1990). However, such excitation-excitation (EE) type cells are but one kind of neural units potentially useful for obtaining binaural information (see, e.g., the introduction and references of Breebaart *et al.*, 2001). With current knowledge, the interaction between the binaural processor and higher level cognitive processes can only be addressed through indirect psychophysical evidence.

For a single source in free field, sound from only one direction arrives at the ears of a listener and thus causally determines the ITD and *interaural level difference* (ILD) cues (Gaik, 1993), which appear in the auditory system as a result of reflections, diffraction, and resonance effects caused by the head, torso, and the external ears of the listener. However, in complex listening situations, i.e., in the presence of several sound sources and/or room reflections, it often occurs that sound from several different directions concurrently reaches the position of the listener. Furthermore, the superposition of sound emanating from several directions results in instantaneous ITD and ILD cues that most of the time do not correspond to any of the source directions. Nevertheless, humans have a remarkable ability to resolve such complex composites of sound into separate localizable auditory events at directions corresponding to the sound sources.

Few binaural models have specifically considered localization in complex listening situations. To begin with, Blauert and Cobben (1978) investigated a model with the essential features of most current models, including a simulation of the auditory periphery and cross-correlation analysis. In a precedence effect experiment they concluded that the correct cross-correlation peaks were available but the model could not explain how to identify them. Later, Lindemann (1986a) extended the model with contralateral and temporal inhibition, combining the analysis of both ITD and ILD cues within a single structure that was shown to be able to simulate several precedence effect phenomena (Lindemann, 1986b). The model of Lindemann was further extended by Gaik (1993) to take into account naturally occurring combinations of ITD and ILD cues in free field. A different phenomenological model, using localization inhibition controlled by an onset detector, was proposed by Zurek (1987), and developed into a cross-correlation implementation by Martin (1997). Hartung and Trahiotis (2001) were able to simulate the precedence effect for pairs of clicks without any inhibition, just taking into account the properties of the peripheral hearing. However, this model was not able to predict the localization of continuous narrow-band noises in a comparison of several models by Braasch and Blauert (2003). The best results were achieved with a combined analysis of ITD cues with the model of Lindemann (1986a) and ILD cues using a modified excitation-inhibition (EI) model (Breebaart *et al.*, 2001) extended with temporal inhibition. For independent localization of concurrent sources with nonsimultaneous onsets, Braasch (2002) has proposed a cross-correlation difference model.

In this paper, we propose a single modeling mechanism to explain various aspects of source localization in complex listening situations. The basic approach is very straightforward: only ITD and ILD cues occurring at time instants when they represent the direction of one of the sources are selected, while other cues are ignored. It will be shown that the *interaural coherence* (IC) can be used as an indicator for these time instants. More specifically, by selecting ITD and ILD cues coinciding with IC cues larger than a certain threshold, one can in many cases obtain a subset of ITD and ILD cues similar to the corresponding cues of each source presented separately in free field. The proposed cue selection method is implemented in the framework of a model that considers a physically and physiologically motivated periph-

eral stage, whereas the remaining parts are analytically motivated. Fairly standard binaural analysis is used to calculate the instantaneous ITD, ILD, and IC cues. The presented simulation results reflect psychophysical data from a number of localization experiments cited earlier, involving independent distracters and precedence effect conditions.

The paper is organized as follows. The binaural model, including the proposed cue selection mechanism, is described in Sec. II. The simulation results are presented in Sec. III with a short discussion of each case related to similar psychophysical studies. Section IV includes a general discussion of the model and results, followed by conclusions in Sec. V.

II. MODEL DESCRIPTION

The model can be divided into three parts: auditory periphery, binaural processor, and higher model stages. In this section, each of the model stages is described in detail, followed by a discussion of the features of the model.

A. Auditory periphery

Transduction of sound from a source to the ears of a listener is realized by filtering the source signals either with head-related transfer functions (HRTFs) or with measured binaural room impulse responses (BRIRs). HRTF filtering simulates the direction dependent influence of the head and outer ears on the ear input signals. BRIRs additionally include the effect of room reflections in an enclosed space. In multisource scenarios, each source signal is first filtered with a pair of HRTFs or BRIRs corresponding to the simulated location of the source, and the resulting ear input signals are summed before the next processing stage.

The effect of the middle ear is typically described as a bandpass filter. However, since this paper is only considering simulations at single critical bands, the frequency weighting effect of the middle ear has been discarded in the model. The frequency analysis of the basilar membrane is simulated by passing the left and right ear signals through a gammatone filterbank (Patterson *et al.* 1995). Each resulting critical band signal is processed using a model of neural transduction as proposed by Bernstein *et al.* (1999). The envelopes of the signals are first compressed by raising them to the power of 0.23. The compressed signals are subjected to half-wave rectification followed by squaring and a fourth order low-pass filtering with a cutoff frequency of 425 Hz. The resulting nerve firing densities at the corresponding left and right ear critical bands are denoted x_1 and x_2. These parts of the model are implemented using the freely available Matlab toolboxes from Slaney (1998) and Akeroyd (2001).

Internal noise is introduced into the model in order to describe the limited accuracy of the auditory system. For this purpose independent Gaussian noise, filtered with the same gammatone filters as the considered critical band signals, is added to each critical band signal before applying the model of neural transduction. The noise is statistically independent for each critical band, as well as for the left and right ears. For the critical band centered at 2 kHz, a sound pressure level (SPL) of 9.4 dB has been chosen according to Breebaart *et al.* (2001) who fitted the level of the noise to describe detection performance near the threshold of hearing. For other critical bands the level is scaled according to the hearing threshold curves (ISO 389, 1975). For the 500 Hz band, an SPL of 14.2 dB is used.

B. Binaural processor

As mentioned in Sec. I, the present study does not make a specific physiological assumption about the binaural processor. The only assumption is that its output signals (e.g., binaural activity patterns) yield information which can be used by the upper stages of the auditory system for discriminating ITD, ILD, and IC. Given this assumption, the proposed model computes the ITD, ILD, and IC directly. Note that here ITD, ILD, and IC are defined with respect to critical band signals after applying the neural transduction.

The ITD and IC are estimated from the normalized cross-correlation function. Given x_1 and x_2 for a specific center frequency f_c, at the index of each sample n, a running normalized cross-correlation function is computed according to

$$\gamma(n,m)=\frac{a_{12}(n,m)}{\sqrt{a_{11}(n,m)a_{22}(n,m)}}, \tag{1}$$

where

$$\begin{aligned}
a_{12}(n,m)&=\alpha x_1(n-\max\{m,0\})x_2(n-\max\{-m,0\})\\
&\quad+(1-\alpha)a_{12}(n-1,m),\\
a_{11}(n,m)&=\alpha x_1(n-\max\{m,0\})x_1(n-\max\{m,0\})\\
&\quad+(1-\alpha)a_{11}(n-1,m),\\
a_{22}(n,m)&=\alpha x_2(n-\max\{-m,0\})x_2(n-\max\{-m,0\})\\
&\quad+(1-\alpha)a_{22}(n-1,m),
\end{aligned}$$

and $\alpha\in[0,1]$ determines the time constant of the exponentially decaying estimation window

$$T=\frac{1}{\alpha f_s}, \tag{2}$$

where f_s denotes the sampling frequency. $\gamma(n,m)$ is evaluated over time lags in the range of $[-1,1]$ ms, i.e., $m/f_s\in[-1,1]$ ms. The ITD (in samples) is estimated as the lag of the maximum of the normalized cross-correlation function,

$$\tau(n)=\arg\max_m \gamma(n,m). \tag{3}$$

Note that the time resolution of the computed ITD is limited by the sampling interval.

The normalization of the cross-correlation function is introduced in order to get an estimate of the IC, defined as the maximum value of the instantaneous normalized cross-correlation function,

$$c_{12}(n)=\max_m \gamma(n,m). \tag{4}$$

This estimate describes the coherence of the left and right ear input signals. In principle, it has a range of [0,1], where 1 occurs for perfectly coherent x_1 and x_2. However, due to the DC offset of the halfwave rectified signals, the values of c_{12}

are typically higher than 0 even for independent (nonzero) x_1 and x_2. Thus, the effective range of the interaural coherence c_{12} is compressed from [0,1] to [a,1] by the neural transduction. The compression is more pronounced (larger a) at high frequencies, where the low pass filtering of the half-wave rectified critical band signals yields signal envelopes with a higher DC offset than in the signal wave forms (Bernstein and Trahiotis, 1996).

The ILD is computed as

$$\Delta L(n) = 10 \log_{10}\left(\frac{L_2(n,\tau(n))}{L_1(n,\tau(n))}\right), \tag{5}$$

where

$$L_1(n,m) = \alpha x_1^2(n-\max\{m,0\}) + (1-\alpha)L_1(n-1,m),$$

$$L_2(n,m) = \alpha x_2^2(n-\max\{-m,0\}) + (1-\alpha)L_2(n-1,m).$$

Note that due to the envelope compression the resulting ILD estimates will be smaller than the level differences between the ear input signals. For coherent ear input signals with a constant level difference, the estimated ILD (in dB) will be 0.23 times that of the physical signals.

The sum of the signal power of x_1 and x_2 that contributes to the estimated ITD, ILD, and IC cues at time index n is

$$p(n) = L_1(n,\tau(n)) + L_2(n,\tau(n)). \tag{6}$$

Choosing the time constant T is a difficult task. Studies of binaural detection actually suggest that the auditory system integrates binaural data using a double-sided window with time constants of both sides in the order of 20–40 ms (e.g., Kollmeier and Gilkey, 1990). However, a double sided window with this large time constant will not be able to simulate the precedence effect, where the localization of a lead sound should not be influenced by a lagging sound after only a few milliseconds. The difference could be explained by assuming that the auditory system responsible for binaural detection further integrates the binaural data originally derived with a better time resolution. In this paper we have chosen to use a single-sided exponential time window with a time constant of 10 ms, in accordance with the time constant of the temporal inhibition of the model of Lindemann (1986a).

C. Higher model stages

A vast amount of information is available to the upper stages of the auditory system through the signals from the auditory periphery. The focus of this study lies only in the analysis of the three interchannel properties between left and right critical band signals that were defined in the preceding section: ITD, ILD, and IC. It is assumed that at each time instant n the information about the values of these three signal properties, $\{\Delta L(n),\tau(n),c_{12}(n)\}$, is available for further processing in the upper stages of the auditory system.

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 τ_ϕ and ΔL_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.

When several independent sources are concurrently active in free field, the resulting cue triplets $\{\Delta L(n),\tau(n),c_{12}(n)\}$ can be classified into two groups: (1) Cues arising at time instants when only one of the sources has power in that critical band. These cues are similar to the free-field cues [direction is represented in $\{\Delta L(n),\tau(n)\}$, and $c_{12}(n)\approx 1$]. (2) Cues arising when multiple sources have non-negligible power in a critical band. In such a case, the pair $\{\Delta L(n),\tau(n)\}$ does not represent the direction of any single source, unless the superposition of the source signals at the ears of the listener incidentally produces similar cues. Furthermore, when the two sources are assumed to be independent, the cues are fluctuating and $c_{12}(n)<1$. These considerations motivate the following method for selecting ITD and ILD cues. Given the set of all cue pairs, $\{\Delta L(n),\tau(n)\}$, only the subset of pairs is considered which occurs simultaneously with an IC larger than a certain threshold, $c_{12}(n) > c_0$. This subset is denoted

$$\{\Delta L(n),\tau(n)|c_{12}(n)>c_0\}. \tag{7}$$

The same cue selection method is applicable for deriving the direction of a source while suppressing the directions of one or more reflections. When the "first wave front" arrives at the ears of a listener, the evoked ITD and ILD cues are similar to the free-field cues of the source, and $c_{12}(n) \approx 1$. As soon as the first reflection from a different direction arrives, the superposition of the source signal and the reflection results in cues that do not resemble the free-field cues of either the source or the reflection. At the same time IC reduces to $c_{12}(n)<1$, since the direct sound and the reflection superimpose as two signal pairs with different ITD and ILD. Thus, IC can be used as an indicator for whether ITD and ILD cues are similar to free-field cues of sources or not, while ignoring cues related to reflections.

For a given c_0 there are several factors determining how frequently $c_{12}(n)>c_0$. In addition to the number, strengths, and directions of the sound sources and room reflections, $c_{12}(n)$ depends on the specific source signals and on the critical band being analyzed. In many cases, the larger the c_0 the more similar the selected cues are to the free-field cues. However, there is a strong motivation to choose c_0 as small as possible while still getting accurate enough ITD and/or ILD cues, because this will lead to the cues being selected more often, and consequently to a larger proportion of the ear input signals contributing to the localization.

It is assumed that the auditory system adapts c_0 for each specific listening situation, i.e., for each scenario with a con-

stant 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 c_0 is adapted relatively slowly in time. In Sec. III B 1, it is also argued that such an adaptive process may be related to the buildup of the precedence effect. All simulations reported in this paper consider only one specific listening situation at a time. Therefore, for each simulation a single constant c_0 is used.

D. Discussion

The physiological feasibility of the cue selection depends on the human sensitivity to changes in interaural correlation. The topic has been investigated by Pollack and Trittipoe (1959a, 1959b), Gabriel and Colburn (1981), Grantham (1982), Koehnke *et al.* (1986), Jain *et al.* (1991), Culling *et al.* (2001), and Boehnke *et al.* (2002). These investigations agree in that the sensitivity is highest for changes from full correlation, whereas the estimates of the corresponding just noticeable differences (JNDs) have a very large variance. For narrow band noise stimuli centered at 500 Hz, the reported JNDs range from 0.0007 (Jain *et al.*, 1991, fringed condition) to 0.13 (Culling *et al.*, 2001) for different listeners and different stimulus conditions. The sensitivity has been generally found to be lower at higher frequencies. However, all the cited studies have measured sensitivity to correlation of the ear input wave forms instead of correlation computed after applying a model of neural transduction. As discussed in Sec. II B, the model of Bernstein *et al.* (1999) reduces the range of IC, indicating overall lower JNDs of IC as defined in this paper. Furthermore, the model has been specifically fitted to yield constant thresholds at different critical bands when applied to prediction of binaural detection based on changes in IC (Bernstein and Trahiotis, 1996). With these considerations it can be concluded that at least the JNDs reported by Gabriel and Colburn (1981), Koehnke *et al.* (1986), and Jain *et al.* (1991) are within the range of precision needed for the simulations in Sec. III.

The auditory system may not actually use a hard IC threshold for selecting or discarding binaural cues. Instead of pure selection, similar processing could be implemented as an IC based weighting of ITD and ILD cues with a slightly smoother transition. However, the simple selection criterion suffices to illustrate the potential of the proposed method, as will be shown in Sec. III. Interestingly, van de Par *et al.* (2001) have argued that the precision needed for normalization of the cross-correlation function is so high that it is unlikely that the auditory system is performing the normalization *per se*. Since normalized cross correlation, nevertheless, describes the perception of IC well, it will be utilized in this paper.

The cue selection can also be seen as a multiple looks approach for localization. Multiple looks have been previously proposed to explain monaural detection and discrimination performance with increasing signal duration (Viemeister and Wakefield, 1991). The idea is that the auditory system has a short term memory of ''looks'' at the signal, which can be accessed and processed selectively. In the case of localization, the looks would consist of momentary ITD, ILD, and IC cues. With an overview of a set of recent cues, ITDs and ILDs corresponding to high IC values could be adaptively selected.

III. SIMULATION RESULTS

As mentioned earlier, it is assumed that in order to perceive an auditory event at a certain direction, the auditory system needs to obtain cues similar to the free-field cues corresponding to a source at that direction. In the following, the proposed cue selection is applied to several stimuli that have been used in previously published psychophysical studies. In all cases both the selected cues as well as all cues prior to the selection are illustrated, and the implied directions are discussed in relation to the literature.

The effectiveness of the proposed cue selection is assessed using a number of statistical measures. The biases of the ITD and ILD cues with respect to the free-field cues τ_ϕ and ΔL_ϕ are defined as

$$\begin{aligned} b_\tau &= |E\{\tau(n)\} - \tau_\phi|, \\ b_{\Delta L} &= |E\{\Delta L(n)\} - \Delta L_\phi|, \end{aligned} \tag{8}$$

respectively, and the corresponding standard deviations are given by

$$\begin{aligned} \sigma_\tau &= \sqrt{E\{(\tau(n) - E\{\tau(n)\})^2\}}, \\ \sigma_{\Delta L} &= \sqrt{E\{(\Delta L(n) - E\{\Delta L(n)\})^2\}}. \end{aligned} \tag{9}$$

The biases and standard deviations are computed considering only the selected cues [Eq. (7)]. When there is more than one source to be discriminated, these measures are estimated separately for each source by grouping the selected cues at each time instant with the source known to have free-field cues closest to their current values.

For many cases, the larger the cue selection threshold c_0, the smaller the bias and standard deviation. The choice of c_0 is a compromise between the similarity of the selected cues to the free-field cues and the proportion of the ear input signals contributing to the resulting localization. The proportion of the signals contributing to the localization is characterized with the fraction of power represented by the selected parts of the signals, given by

$$p_0 = \frac{E\{p(n)w(n)\}}{E\{p(n)\}}, \tag{10}$$

where $p(n)$ is defined in Eq. (6) and the weighting function $w(n)$ is

$$w(n) = \begin{cases} 1, & \text{if } c_{12}(n) > c_0, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

In this paper, the cue selection is only considered independently at single critical bands. Except for different values of c_0, the typical behavior appears to be fairly similar at critical bands with different center frequencies. For most simulations, we have chosen to use the critical bands centered at 500 Hz and/or 2 kHz. At 500 Hz the binaural processor operates on the input wave forms, whereas at 2 kHz the model of auditory periphery extracts the envelopes of the input signals and feeds them to the binaural processor. Where

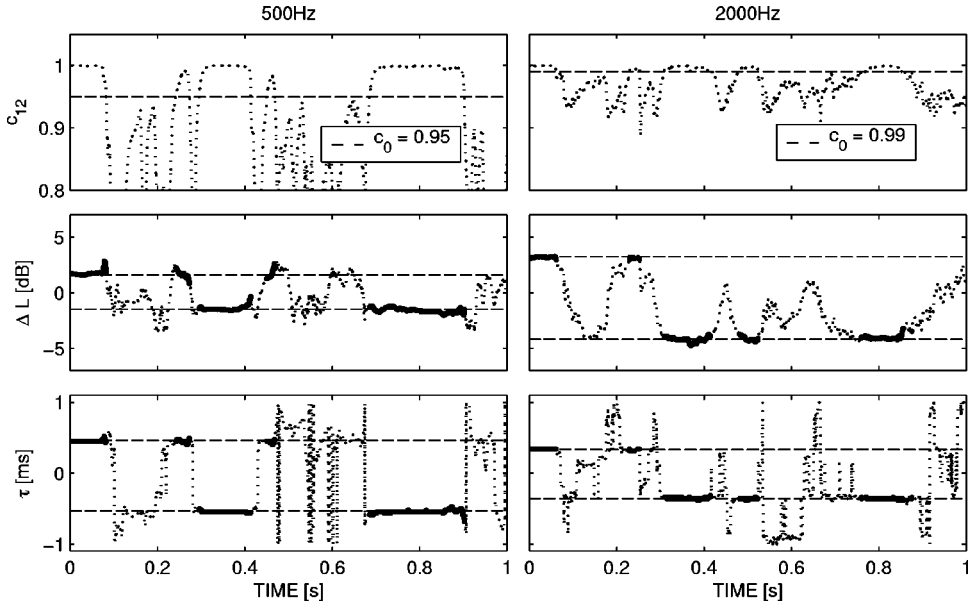

FIG. 2. IC, ILD, and ITD as a function of time for two independent speech sources at ±40° azimuth. Left column, 500 Hz; and right column, 2 kHz critical band. The cue selection thresholds (top row) and the free-field cues of the sources (middle and bottom rows) are indicated with dashed lines. Selected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

appropriate, results for other critical bands are also shown or briefly discussed. However, considering the way the auditory system eventually combines information from different critical bands is beyond the scope of this paper. As mentioned earlier, the simulations are carried out with a single constant cue selection threshold c_0 for each case. It is assumed that the auditory system has already adapted c_0 to be effective for the specific listening situation. Unless otherwise noted, the specific c_0 was chosen such that a visual inspection of the simulation results implies an effective cue selection.

Two kinds of plots are used to illustrate the cue selection. In some cases the instantaneous ITD and ILD values are plotted as a function of time, marking the values which are selected. For other examples, the effect of the cue selection is visualized by plotting short-time estimates of *probability density functions* (PDFs) of the selected ITD and ILD cues. Unless otherwise noted, the PDFs are estimated by computing histograms of ITD and ILD cues for a time span of 1.6 s. The height of the maximum peak is normalized to one in all PDFs. In both types of plots, free-field cues resulting from simulations of the same source signals without concurrent sound sources or reflections, are also indicated (the Matlab code used for these simulations is available at http://www.acoustics.hut.fi/software/cueselection/).

Listening situations in free field are simulated using HRTFs measured with the KEMAR dummy head with large pinnae, taken from the CIPIC HRTF Database (Algazi *et al.*, 2001). All simulated sound sources are located in the frontal horizontal plane, and, unless otherwise noted, all the stimuli are aligned to 60 dB SPL averaged over the whole stimulus length.

A. Independent sources in free-field

In this section, the cue selection method is applied to independent stimuli in an anechoic environment. As the first example, the operation of the selection procedure is illustrated in detail for the case of independent speech sources at different directions. Subsequently, simulation results of the effect of target-to-distracter ratio (T/D) on localization of the target stimulus are presented.

1. Concurrent speech

Localization of a speech target in the presence of one or more competing speech sources has been investigated by Hawley *et al.* (1999) and Drullman and Bronkhorst (2000). Drullman and Bronkhorst (2000) utilized an anechoic virtual environment using both individualized and nonindividualized HRTFs for binaural reproduction of the stimuli. They reported slight but statistically significant degradation in localization performance when the number of competing talkers was increased beyond 2. The experiment of Hawley *et al.* (1999), on the other hand, was conducted in a ''sound-field room'' (reverberation time of approximately 200 ms), as well as using headphone reproduction of the stimuli recorded binaurally in the same room. While not strictly anechoic, their results are also useful for evaluating our anechoic simulation results. Hawley *et al.* (1999) found that apart from occasional confusions between the target and the distracters, increasing the number of competitors from 1 to 3 had no significant effect on localization accuracy. As discussed in Sec. I, room reflections generally make the localization task more difficult, so a similar or a better result would be expected to occur in an anechoic situation. Note that the overall localization performance reported by Drullman and Bronkhorst (2000) was fairly poor, and the results may have been affected by a relatively complex task requiring listeners to recognize the target talker prior to judging its location.

Based on the previous discussion, the cue selection has to yield ITD and ILD cues similar to the free-field cues of each of the speech sources in order to correctly predict the directions of the perceived auditory events. Three simulations were carried out with 2, 3, and 5 concurrent speech sources. The signal of each source consisted of a different phonetically balanced sentence from the Harvard IEEE list (IEEE, 1969) recorded by the same male speaker. As the first

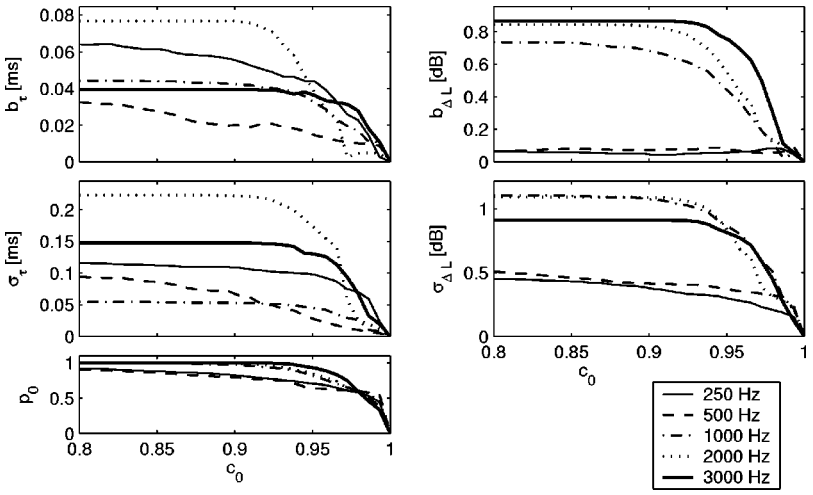

FIG. 3. ITD and ILD bias (top panels), standard deviation (middle panels), and relative power (bottom left panel) of the selected signal portions as a function of the cue selection threshold c_0 for two independent speech sources. Data are shown for the 250, 500, 1000, 2000, and 3000 Hz critical bands.

case, 2 speech sources were simulated at azimuthal angles of $\pm 40°$. Figure 2 shows the IC, ILD, and ITD as a function of time for the critical bands with center frequencies of 500 Hz and 2 kHz. The free-field cues which would occur with a separate simulation of the sources at the same angles are indicated with the dashed lines. The selected ITD and ILD cues [Eq. (7)] are marked with bold solid lines. Thresholds of $c_0 = 0.95$ and $c_0 = 0.99$ were used for the 500 Hz and 2 kHz critical bands, respectively, resulting in 65% and 54% selected signal power [Eq. (10)]. The selected cues are always close to the free-field cues, implying perception of two auditory events located at the directions of the sources, as reported in the literature. As expected, due to the neural transduction IC has a smaller range at the 2 kHz critical band than at the 500 Hz critical band. Consequently, a larger c_0 is required.

The performance of the cue selection was assessed as a function of c_0 for the same two speech sources and the critical bands with center frequencies of 250, 500, 1000, 2000, and 3000 Hz. Figure 3 shows the ITD and ILD biases [Eq. (8)] and standard deviations [Eq. (9)], as well as the fraction of signal power corresponding to the selected cues [Eq. (10)] as a function of c_0. The biases and standard deviations were computed for both sources separately, as described earlier, and then averaged over 1.6 s of the signals. The graphs indicate that both the biases and the standard deviations decrease with increasing c_0. Thus, the larger the c_0, the closer the obtained cues are to the reference free-field values. Furthermore, the selected signal power decreases gradually until fairly high values of c_0. The general trend of having higher absolute ILD errors at high frequencies is related to the overall larger range of ILDs occuring at high frequencies due to more efficient head shadowing.

The simulation with three independent talkers was performed with speech sources at 0° and $\pm 30°$ azimuth, and the simulation of five talkers with two additional sources at $\pm 80°$ azimuth. In both cases the results were fairly similar at different critical bands, so the data are only shown for the 500 Hz band. Panels (A) and (B) of Fig. 4 show PDFs of ITD and ILD without the cue selection for the three and five

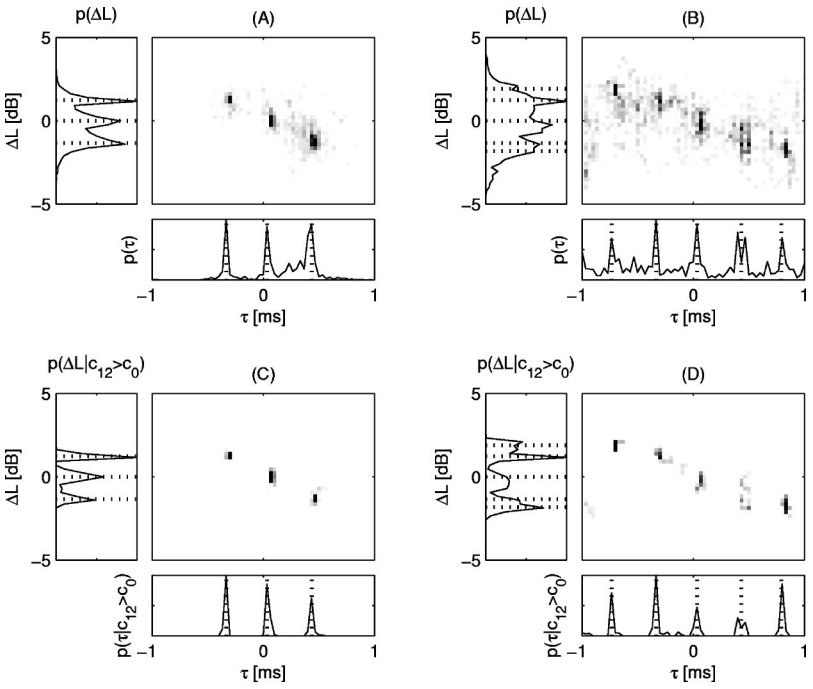

FIG. 4. PDFs of ITD and ILD for three (A) and five (B) independent speech sources and corresponding PDFs when cue selection is applied [(C) and (D)]. The values of the free-field cues for each source are indicated with dotted lines. Data are shown for the 500 Hz critical band.

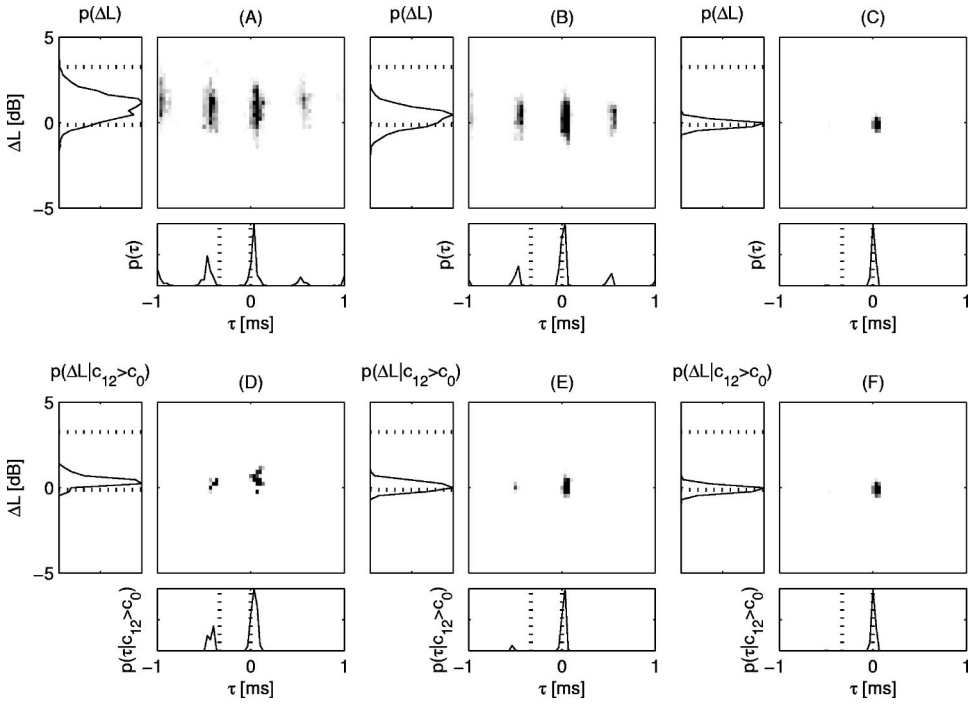

FIG. 5. PDFs of ITD and ILD for a click-train and white Gaussian noise at different T/D ratios: −3, −9, −21 dB (A)–(C), and the corresponding PDFs when cue selection is applied (D)–(F). The values of the free-field cues are indicated with dotted lines. Data are shown for the 2 kHz critical band.

speech sources, respectively. Panels (C) and (D) of Fig. 4 show similar PDFs of the selected cues. The selection threshold was set at $c_0 = 0.99$ corresponding to 54% selected signal power for the three sources and 22% for the five sources. In both cases, even the PDFs considering all cues show ITD peaks at approximately correct locations, and the cue selection can be seen to enhance the peaks. With the cue selection, the widths of the peaks (i.e., the standard deviations of ITD and ILD) in the three source case are as narrow as in separate one source free-field simulations, which implies robust localization of three auditory events corresponding to the psychophysical results of Hawley *et al.* (1999) and Drullman and Bronkhorst (2000). In the case of five sources, the peaks get slightly broader. The ITD peaks are still narrow and correctly located but at the 500 Hz critical band, the range of ILD cues is insufficient for distinct peaks to appear along the ILD axis. This result is also in line with the classic duplex theory (Rayleigh, 1907) of sound localization, stating that at low frequencies ITD cues are more salient than ILD cues.

2. Click-train and noise

Good and Gilkey (1996) and Good *et al.* (1997) studied the localization of a click-train target in the presence of a simultaneous noise distracter. Using loudspeaker reproduction in an anechoic chamber, localization performance was shown to degrade monotonously with a decreasing target-to-distracter ratio (T/D). The investigated T/D ratios were defined relative to the individual detection threshold of each listener for the case when the target sound was presented from the same direction as the distracter. With a target level just a few dB above the detection threshold, localization performance in the left-right direction (e.g., frontal horizontal plane) was still found to be nearly as good as without the distracter. The degradation started earlier and was more severe for the up-down and front-back directions. The results for the left-right direction were later confirmed by Lorenzi *et al.* (1999), who conducted a similar experiment with sound sources in the frontal horizontal plane. However, the detection levels of Lorenzi *et al.* (1999) were slightly higher, maybe due to utilization of a sound-treated chamber instead of a strictly anechoic environment. Furthermore, Lorenzi *et al.* (1999) found a degradation in performance when the stimuli were low-pass filtered at 1.6 kHz, unlike when the stimuli were high pass filtered at the same frequency.

A simulation was carried out with a white noise distracter directly in front of the listener and a click-train target with a rate of 100 Hz located at 30° azimuth. Assuming a detection level of −11 dB (the highest value in Good *et al.* 1997), the chosen absolute T/D of −3, −9, and −21 dB correspond to the relative T/D of 8, 2, and −10 dB, respectively, as investigated by Good and Gilkey (1996). The PDFs for the critical band centered at 500 Hz did not yield a clear peak corresponding to the direction of the click train. Motivated by the fact that in this case higher frequencies are more important for directional discrimination (Lorenzi *et al.*, 1999), we investigated further the 2 kHz critical band. Panels (A)–(C) of Fig. 5 show PDFs of ITD and ILD without the cue selection for the selected T/D ratios. Corresponding PDFs obtained by the cue selection [Eq. (7)] are shown in panels (D)–(F). The thresholds for the panels (D)–(F) were $c_0 = 0.990$, $c_0 = 0.992$, and $c_0 = 0.992$, respectively, resulting in 3%, 9%, and 99% of the signal power being represented by the selected cues.

The PDFs in Fig. 5 imply that the target is localized as a separate auditory event for the T/D ratios of −3 dB and −9 dB. However, for the lowest T/D ratio the target click-train is no longer individually localizable, as also suggested by the results of Good and Gilkey (1996). In panels (A) and (B), ITD peaks are seen to rise at regular intervals due to the periodicity of the cross-correlation function, while the cue selection suppresses the periodical peaks as shown in panels

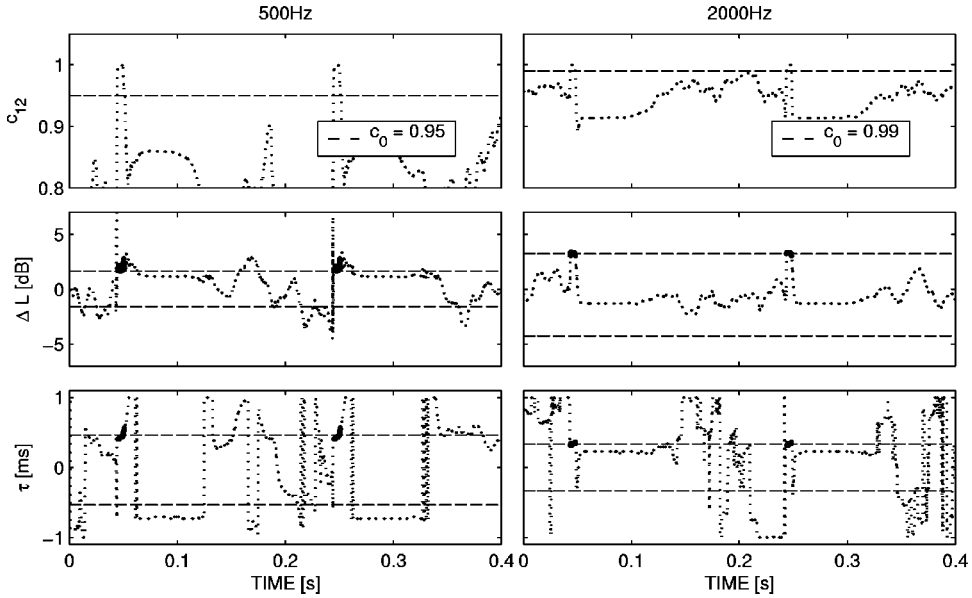

FIG. 6. IC, ILD, and ITD as a function of time for a lead/lag click-train with a rate of 5 Hz and an ICI of 5 ms. Left column, 500 Hz; and right column, 2 kHz critical band. The cue selection thresholds (top row) and the free-field cues of the sources (middle and bottom rows) are indicated with dashed lines. Selected cues are marked with bold solid lines.

(D) and (E). Note that when the click-train is individually localizable, only the recovered ITD cues are close to the free-field cues of both sources, whereas a single broad ILD peak appears. This is in line with the findings of Braasch (2003) that in the presence of a distracter, ILDs are less reliable cues for localization, and that ITDs also gain more importance in the subjective localization judgment. The ITD peaks corresponding to the click-train are also shifted away from the distracter. Such a pushing effect caused by a distracter in front of the listener was observed for one listener in a similar experiment (Lorenzi *et al.*, 1999) and for most listeners when the target was an independent noise signal (Braasch and Hartung 2002). On the contrary, Good and Gilkey (1996) reported a pulling effect, which was also the case for two listeners in the experiment of Lorenzi *et al.* (1999).

B. Precedence effect

This section illustrates the cue selection within the context of the precedence effect. Pairs of clicks are used to demonstrate the results for wide band signals (in this case a signal with at least the width of a critical band). Sinusoidal tones are simulated with different onset rates and the cues obtained during the onset are shown to agree with results reported in the literature.

1. Pairs of clicks

In a classical precedence effect experiment, a lead/lag pair of clicks is presented to the listener (Blauert, 1997; Litovsky *et al.*, 1999). The leading click is first emitted from one direction, followed by another identical click from another direction after an *interclick interval* (ICI) of a few milliseconds. As discussed in Sec. I, the directional perception changes depending on ICI.

Figure 6 shows IC, ILD, and ITD as a function of time for a click train with a rate of 5 Hz analyzed at the critical bands centered at 500 Hz and 2 kHz. The lead source is simulated at 40° and the lag at −40° azimuth with an ICI of 5 ms. As expected based on earlier discussion, IC is close to one whenever only the lead sound is within the analysis time window. As soon as the lag reaches the ears of the listener, the superposition of the two clicks reduces the IC. The cues obtained by the selection with $c_0=0.95$ for the 500 Hz and $c_0=0.985$ for the 2 kHz critical band are shown in the figure, and the free-field cues of both sources are indicated with dashed lines. The selected cues are close to the free-field cues of the leading source and the cues related to the lag are ignored, as is known to happen based on psychophysical studies (Litovsky *et al.* 1999). The fluctuation in the cues before each new click pair is due to the internal noise of the model.

The performance of the cue selection was again assessed as a function of c_0 for the critical bands with center frequencies of 250, 500, 1000, 2000, and 3000 Hz. The statistical measures were calculated from a 1.6 s signal segment. Figure 7 shows ITD and ILD biases [Eq. (8)] and standard deviations [Eq. (9)], as well as the power of the selected cues [Eq. (10)] as a function of c_0. The biases and standard deviations were computed related to the free-field cues of the leading source, since localization of the lag should be suppressed if the selection works correctly. Both the biases and standard deviations decrease as c_0 increases. Thus the larger the cue selection threshold c_0, the more similar the selected cues are to the free-field cues of the leading source.

At a single critical band, the energy of the clicks is spread over time due to the gammatone filtering and the model of neural transduction. Therefore, with an ICI of 5 ms, a large proportion of the critical band signals related to the clicks of a pair is overlapping, and only a small part of the energy of the lead click appears in the critical band signals before the lag. Consequently, the relative signal power corresponding to the selected cues is fairly low when requiring small bias and standard deviation, as can be seen in the left bottom panel of Fig. 7.

Localization as a function of ICI: The previous experiment was repeated for ICIs in the range of 0–20 ms using the 500 Hz critical band. The chosen range of delays includes summing localization, localization suppression, and independent localization of both clicks without the precedence effect

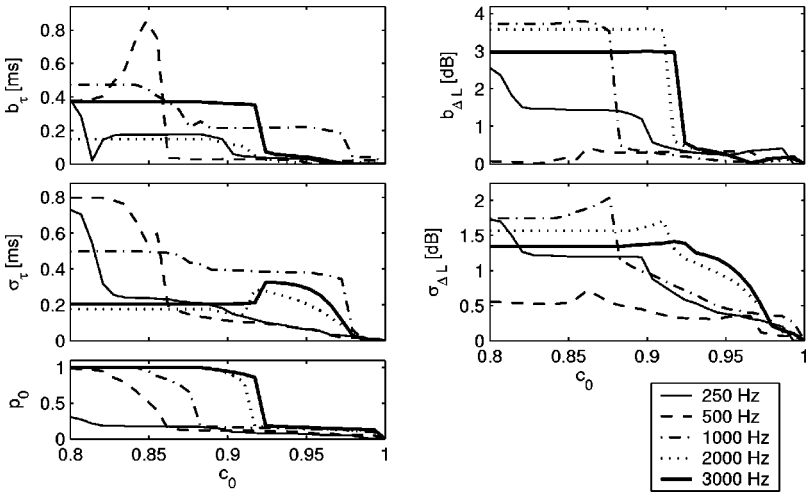

FIG. 7. ITD and ILD bias, standard deviation, and relative power of the selected signal portions as a function of the cue selection threshold c_0 for a lead/lag click-train. Data are shown for the 250, 500, 1000, 2000, and 3000 Hz critical bands.

(Litovsky *et al.*, 1999). For all previous simulations, a suitable c_0 was chosen as a compromise between similarity of the cues to free-field cues and how frequently cues are selected. Here, each ICI corresponds to a different listening situation, since the different delays of the lag imply different acoustical environments. It is thus expected that the most effective c_0 may also differ depending on ICI.

Several different criteria for determining c_0 were assessed. Indeed, using the same c_0 for all ICIs did not yield the desired results. The criterion of adapting c_0 such that the relative power of the selected cues [Eq. (10)] had the same value for each simulation did not yield good results either. Thus, a third criterion was adopted. The cue selection threshold c_0 was determined numerically for each simulation such that σ_τ (the narrowness of the peaks in the PDFs of ITD) was equal to 15 μs. This could be explained with a hypothetical auditory mechanism adapting c_0 in time with the aim of making ITD and/or ILD standard deviation sufficiently small. Small standard deviations indicate small fluctuations of the selected cues in time and thus non-time-varying localization of auditory events. The resulting PDFs of ITD and ILD as a function of ICI with and without the cue selection are shown in Fig. 8.

The PDFs without the cue selection (rows 1 and 2 in Fig. 8) indicate two independently localized auditory events for most ICIs above 1 ms. Furthermore, the predicted directions depend strongly on the delay. On the contrary, the PDFs with the cue selection show that the selected cues correctly predict all the three phases of the precedence effect (summing localization, localization suppression, and independent localization). At delays less than approximately 1 ms the ITD peak moves to the side as the delay increases, as desired, but the ILD cues do not indicate the same direction as the ITD cues. However, this is also in line with existing psychophysical literature. Anomalies of the precedence effect have been observed in listening tests with band pass filtered clicks

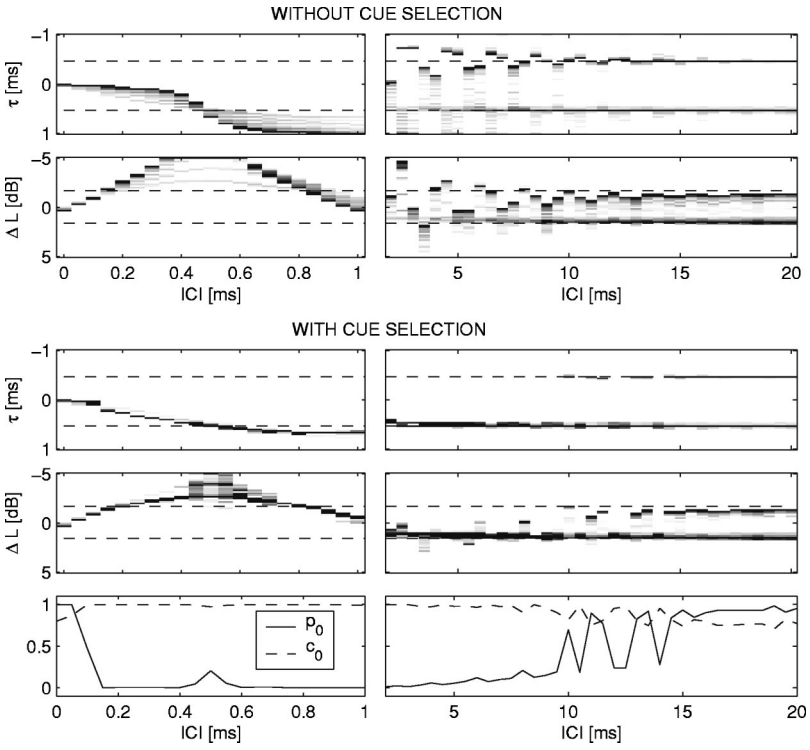

FIG. 8. PDFs of ITD and ILD as a function of the interclick interval (ICI) for a click-train: without cue selection (rows 1 and 2) and with cue selection (rows 3 and 4). Cue selection threshold c_0 and relative power p_0 of the selected signal portion (bottom row).

(Blauert and Cobben, 1978), suggesting a contribution of the extracted misleading ILDs to the localization judgment. For delays within the range of approximately 1–10 ms there is only one significant peak in the PDFs, indicating localization in the direction of the lead. For larger delays two peaks appear, suggesting two independently localized auditory events. Note that the fusion of two clicks has been found to sometimes break down earlier, but 10 ms is within the range of reported critical thresholds for localization dominance (Litovsky *et al.*, 1999; Litovsky and Shinn-Cunningham, 2001).

The bottom row of Fig. 8 shows the selection threshold c_0 and the relative power p_0 of the signal corresponding to the selected cues as a function of the ICI. For most ICIs up to approximately 8 ms, the relative power of the selected signal portion almost vanishes. However, there is one characteristic peak of p_0 at approximately 0.5 ms. The experiment was repeated for a number of critical bands in the range of 400 to 600 Hz with the observation that the location of this peak moves along the ICI axis as a function of the center frequency of the considered critical band. Furthermore, the general trends of the selected cues were very similar to those at the 500 Hz band in that they all strongly implied the three phases of the precedence effect. Thus, by considering a number of critical bands the three phases of the precedence effect can indeed be explained by the cue selection such that at each ICI a signal portion with nonvanishing power is selected.

Cue selection threshold and precedence buildup: For the previous experiment, it was hypothesized that the criterion for determining c_0 is the standard deviation of ITD and/or ILD. The computation of these quantities involves determining the number of peaks (i.e., the number of individually localized auditory events) adaptively in time, which might be related to the buildup of precedence. A buildup occurs when a lead/lag stimulus with ICI close to the echo threshold is repeated several times. During the first few stimulus pairs, the precedence effect is not active and two auditory events are independently perceived. After the buildup, the clicks merge to a single auditory event in the direction of the lead (Freyman *et al.*, 1991). An adaptive process determining c_0 would require a certain amount of stimulus activity and time until an effective c_0 is determined and it could thus explain the time-varying operation of the precedence effect.

The precedence effect literature also discusses a breakdown of precedence when, for instance, the directions of the lead and lag are suddenly swapped (Clifton, 1987; Blauert, 1997; Litovsky *et al.*, 1999). However, more recent results of Djelani and Blauert (2001, 2002) indicate that the buildup is direction specific, suggesting further that what has been earlier reported as breakdown of precedence is rather a consequence of precedence not being built up for a new lag direction. Djelani and Blauert (2002) also showed that without stimulus activity the effect of the buildup decays slowly by itself, which supports the idea of an adaptive c_0. In order to model the direction-specific buildup, c_0 would also need to be defined as a function of direction. However, testing and developing the corresponding adaptation method is beyond the scope of this paper and will be part of the future work.

2. Onset rate of a sinusoidal tone

Rakerd and Hartmann (1986) investigated the effect of the onset time of a 500 Hz sinusoidal tone on localization in the presence of a single reflection. In the case of a sinusoidal tone, the steady state ITD and ILD cues result from the coherent sum of the direct and reflected sound at the ears of a listener. Often these cues do not imply the direction of either the direct sound or the reflection. Rakerd and Hartmann (1986) found that the onset rate of the tone was a critical factor in determining how much the misleading steady state cues contributed to the localization judgment of human listeners. For fast onsets, localization was based on the correct onset cues, unlike when the level of the tone raised slowly. The cue selection cannot, as such, explain the discounting of the steady state cues, which always have IC close to one. However, considering just the onsets the following results reflect the psychophysical findings of Rakerd and Hartmann (1986).

Figure 9 shows IC, ILD, and ITD as a function of time for a 500 Hz tone with onset times of 0, 5, and 50 ms. The simulated case corresponds approximately to the ''WDB room'' and ''reflection source 6'' condition reported by Rakerd and Hartmann (1986). The direct sound is simulated in front of the listener, and the reflection arrives with a delay of 1.4 ms from an azimuthal angle of 30°. A linear onset ramp is used and the steady state level of the tone is set to 65 dB SPL. The ITD and ILD cues selected with a threshold of $c_0=0.93$ are marked with bold solid lines and the free-field cues of the direct sound and the reflection are indicated with dashed lines. Note that the direct sound reaches the ears of the listener at approximately 7 ms. For onset times of 0 and 5 ms, ITD and ILD cues are similar to the free-field cues at the time when IC reaches the threshold. However, with an onset time of 50 ms the ITD and ILD cues no longer correspond to the free-field cues, which is suggested by the degraded localization performance in the experiment of Rakerd and Hartmann (1986).

In order to predict the final localization judgment, another selection mechanism would be needed to only include the localization cues at the time instants when the cue selection becomes effective. The dependence on the onset rate can be explained by considering the input signals of the binaural processor. During the onset, the level of the reflected sound follows that of the direct sound with a delay of 1.4 ms. Thus, the slower the onset, the smaller the difference. The critical moment is when the level of the direct sound rises high enough above the level of the internal noise to yield IC above the selection threshold. If the reflection has nonnegligible power at that time, localization cues will be biased to the steady state direction already when the selection begins.

C. Independent sources in a reverberant environment

As a final test for the model, the localization of 1 and 2 speech sources was simulated in a reverberant environment. The utilized BRIRs were measured with a Neumann KU 80 dummy head in an empty lecture hall with reverberation times of 2.0 and 1.4 s at the octave bands centered at 500 and

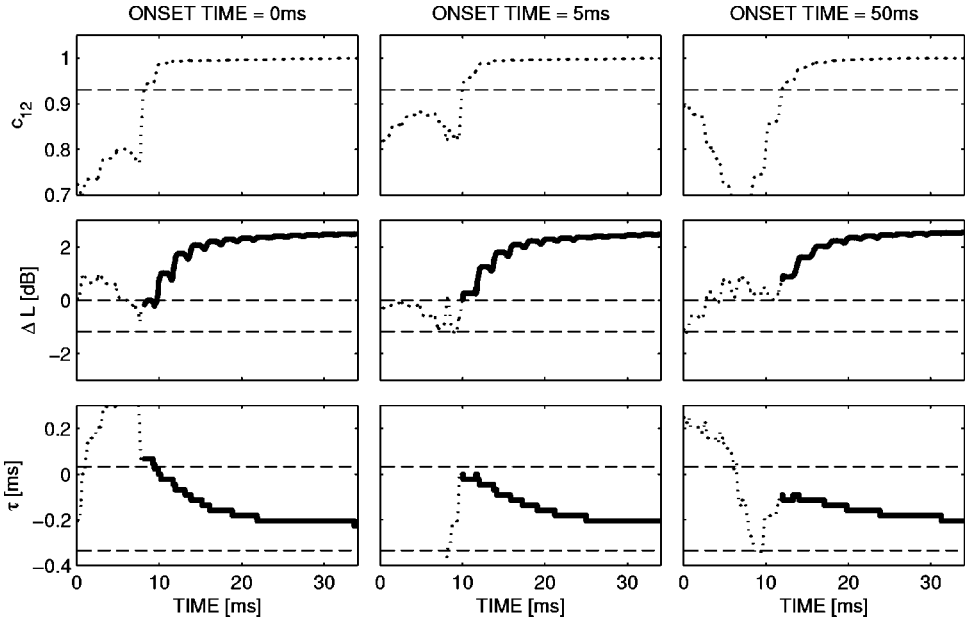

FIG. 9. IC, ILD, and ITD as a function of time for a 500 Hz sinusoidal tone and one reflection. The columns from left to right show results for onset times of 0 ms, 5 ms, and 50 ms. The cue selection threshold of $c_0=0.95$ (top row) and the free-field cues of the source and the reflection (middle and bottom rows) are indicated with dashed lines. Selected cues are marked with bold solid lines. Data are shown for the 500 Hz critical band.

2000 Hz, respectively. The same phonetically balanced speech samples as used in Sec. III A 1 were convolved with BRIRs simulating sources at 30° azimuth for the case of one source and ±30° for the two sources. The case of two talkers included again two different sentences uttered by the same male speaker. For computing the free-field cues, the BRIRs were truncated to 2.3 ms, such that the effect of the reflections was ignored.

The chosen hall is a very difficult case for localization due to lots of diffuse reflections from the tables and benches all around the simulated listening position. At the 500 Hz critical band, the ITD and ILD cues prior to the selection did not yield any meaningful data for localization. The cue selection resulted in high peaks close to the free-field cues, but it was not able to suppress all other peaks implying different directions. A subsequent investigation showed that these erroneous peaks appear at different locations at different critical bands. Thus, processing of localization information across critical bands should be able to further suppress them. At 2 kHz, the results for a single critical band were clearer and they will be illustrated here.

Panels (A) and (B) of Fig. 10 show PDFs of ITD and ILD without the cue selection, and panels (C) and (D) show the corresponding PDFs of the selected cues. Since the cue selection in this case samples the ITD and ILD relatively infrequently, the PDFs were computed considering 3 s of signal. Similar results are obtained when the PDFs are computed from different time intervals. The cue selection criterion for both the 1 and 2 source scenarios was $c_0=0.99$, resulting in 1% of the signal power corresponding to the selected cues. Without the cue selection, the PDFs do not yield much information for localization in either of the cases.

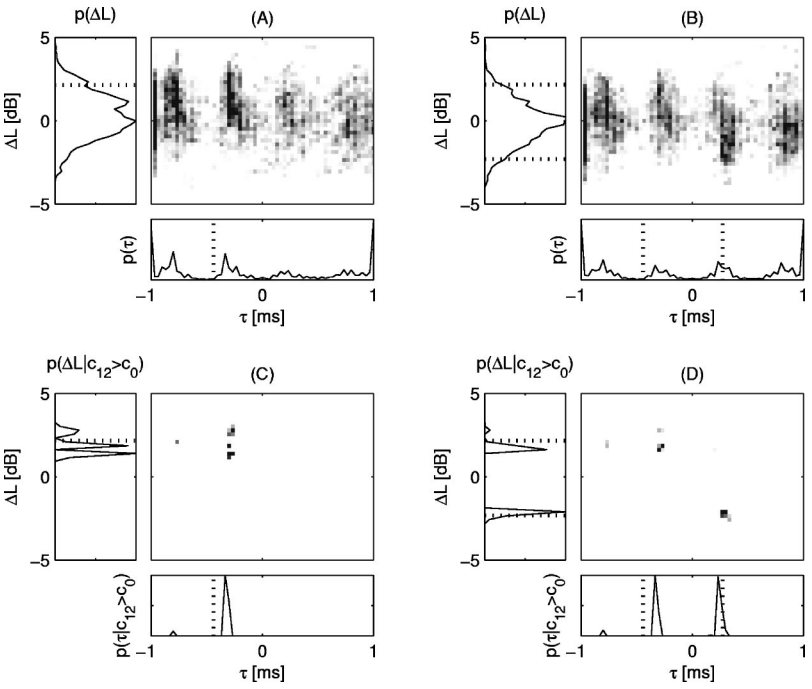

FIG. 10. PDFs of ITD and ILD for 1 (A) and 2 (B) speech sources in a reverberant hall and the corresponding PDFs when cue selection is applied (C) and (D). The values of the free-field cues for each source are indicated with dotted lines. Data are shown for the 2 kHz critical band.

Periodicity of the cross-correlation function is clearly visible and it is difficult to distinguish between the one and two source cases. However, with the cue selection sharp peaks arise relatively close to the free-field cues. In the two source case, the right source is practically correctly localized, whereas the ITD cues of the left source are slightly biased towards the center. Note that contrary to the results in Sec. III A 2, the localization is in this case shifted towards the competing sound source. As discussed, also this kind of a pulling effect has been reported in psychoacoustical studies (Good and Gilkey, 1996; Lorenzi *et al.*, 1999; Braasch and Hartung, 2002).

IV. GENERAL DISCUSSION

In the preceding sections, the selection of ITD and ILD cues based on IC was introduced into a localization model and applied to simulations of a number of complex listening scenarios. In comparison to several existing localization models, a significant difference in the proposed method is the way that the signal power at each time instant affects the localization judgment. In models not designed for complex listening situations, the localization cues and subsequently the final localization judgment are often derived from a time window including the whole stimulus, or of a time integration of a binaural activity pattern computed with running non-normalized cross correlation. In such cases, the contribution of each time instant to the final localization depends on the instantaneous power. In our approach, only the cues during the selected time instants contribute to localization. Thus the model can in many cases neglect localization information corresponding to time instants with high power, if the power is high due to concurrent activity of several sound sources (or concurrent activity of sources and reflections). The relative power of individual sources also affects how often ITD and ILD cues corresponding to each source are selected.

The proposed model also bears resemblance to earlier models of the precedence effect. The temporal inhibition of the model of Lindemann (1986a) tends to hold the highest peaks of the running cross-correlation function (calculated with the stationary inhibition that incorporates ILDs into the model). The higher a peak (i.e., the higher the IC at the corresponding time instant), the stronger the temporal inhibition. The cue selection achieves a somewhat similar effect without a need for an explicit temporal inhibition mechanism, since the localization suppression is directly related to the IC estimated with a similar time window. However, the effect can also be quite different in some scenarios. Whereas the model of Lindemann (1986a) only ''remembers'' the peaks corresponding to a high IC for a short time (time constant of 10 ms), the cue selection with a slowly varying c_0 has a much longer memory. The frequency of the time instants when the direct sound of only one source dominates within a critical band depends on the complexity of the listening situation. In complex cases (e.g., Sec. III C), only a small fraction of the ear input signals contribute to localization, and new localization information may be acquired relatively infrequently. We, nevertheless, assume that it is the cues at these instants of time that determine the source localization. During the time when no cues are selected, the localization of the corresponding auditory events is assumed to be determined by the previously selected cues, which is in principle possible. Localization of sinusoidal tones based only on their onsets (Rakerd and Hartmann 1985, 1986) and a related demonstration called the ''Franssen effect'' (Franssen, 1960; Hartmann and Rakerd, 1989) show that a derived localization judgment can persist for several seconds after the related localization cues have occurred. In precedence effect conditions (Sec. III B) the cue selection naturally derives most localization information from signal onsets, as is explicitly done in the model of Zurek (1987) (see also Martin, 1997). However, the cue selection is not limited to getting information from onsets only, and it does not necessarily include all onsets.

Throughout the paper, the resulting ITD and ILD cues were considered separately instead of deriving a combined localization judgment. The mutual role of ITDs and ILDs is often characterized with time-intensity trading ratios (Blauert, 1997) or in the form of the classic duplex theory (Rayleigh, 1907): ITD cues dominate localization at low frequencies and ILDs at high frequencies. However, in complex listening situations the relative weights of these cues may change. Wightman and Kistler (1992) have shown that in the presence of conflicting ITD and ILD cues, ITD cues will dominate the localization judgment of broadband noise as long as low frequency energy is present. Furthermore, Braasch (2003) has found that the presence of a distracting sound source even strengthens the weight of ITD cues. Nevertheless, the results of Rakerd and Hartmann (1986) suggest that steady-state ITDs can sometimes be completely neglected, unlike ILD cues. Considering the relative weights of ITD and ILD cues in more detail is beyond the scope of this paper. However, in future work it will be interesting to assess whether the proposed cue selection reflects the relative importance of ITD and ILD cues, i.e., whether the cue selection, for example, recovers more reliably ITD cues in cases where they are weighted more than ILD cues, and vice versa.

The cue selection mechanism could be seen to perform a function that Litovsky and Shinn-Cunningham (2001) have characterized as ''a general process that enables robust localization not only in the presence of echoes, but whenever any competing information from a second source arrives before the direction of a previous source has been computed.'' For the purposes of this paper, ITD and IC cues were analyzed using a cross-correlation model, whereas ILDs were computed independently. Similar cue selection could also be implemented in other localization models, such as the excitation-inhibition (EI) model of Breebaart *et al.* (2001) involving joint analysis of ITD and ILD cues within a physiologically motivated structure. In the EI model, full coherence is not represented by maximum activity but by zero activity. Thus, as opposed to specifying a lower bound of IC for the cue selection, an upper bound of activity would need to be determined.

As shown in Sec. III, the cue selection model was able to simulate most psychophysical results reviewed in the introduction by using a selection threshold adapted to each specific listening scenario. Although this paper is limited to

localization based on binaural cues, it should be mentioned that the precedence effect has also been observed in the median sagittal plane where the localization is based on spectral cues instead of interaural differences (Blauert, 1971; Litovsky *et al.,* 1997). Thus, the cue selection model does not fully describe the operation of the precedence effect. Furthermore, the model cannot as such explain the discounting of ITD and ILD cues occurring simultaneously with a high IC during the steady state sound in two scenarios: a sinusoidal tone presented in a room (Rakerd and Hartmann 1985, 1986; Hartmann and Rakerd, 1989) and two independent noise sources with the same envelopes presented from different directions (Braasch, 2002). The psychophysical results of Litovsky *et al.* (1997) show that the localization suppression is somewhat weaker in the median plane than in the horizontal plane, which could be interpreted as evidence for another suppression mechanism, possibly operating simultaneously with a binaural mechanism such as the proposed cue selection. Indeed, simulating all the results cited in this paragraph would appear to require some additional form of temporal inhibition.

V. CONCLUSIONS

A cue selection mechanism was presented for modeling source localization in complex listening scenarios. The cue selection can simulate both localization of several concurrently active independent sources and suppression of the localization of reflected sound by considering ITD and ILD cues only when IC at the corresponding critical band is larger than a certain threshold. It was shown that at time instants when this occurs, ITD and ILD are likely to represent the direction of one of the sources. Thus, by looking at the different ITD and ILD values during the selected time instants one can obtain information about the direction of each source.

The proposed cue selection mechanism was implemented in the framework of a binaural model considering the known periphery of the auditory system. The remaining parts of the model were analytically motivated for the sake of focus on the cue selection method without having to consider the specific properties and limitations of existing physiologically motivated models. Nevertheless, it was pointed out in the discussion that in principle the proposed cue selection method is physiologically feasible.

The binaural model with the proposed cue selection was verified with the results of a number of psychophysical studies from the literature. The simulation results suggest relatively reliable localization of concurrent speech sources both in anechoic and reverberant environments. The effect of target-to-distracter ratio corresponds qualitatively to published results of localization of a click-train in the presence of a noise distracter. Localization dominance is correctly reproduced for click pairs and for the onsets of sinusoidal tones. It was also hypothesized that the buildup of precedence may be related to the time the auditory system needs to find a cue selection threshold which is effective for the specific listening situation. As a final test, the model was applied for source localization in a reverberant hall with one and two speech sources. The results suggest that also in this most complex case the model is able to obtain cues corresponding to the directions of the sources.

ACKNOWLEDGMENTS

The authors gratefully acknowledge discussions with Frank Baumgarte, Jens Blauert, Toni Hirvonen, Aki Härmä, Matti Karjalainen, Kalle Palomäki, and John Worley. The work of Juha Merimaa has been supported by the research training network for Hearing Organization and Recognition of Speech in Europe (HOARSE, HPRN-CT-2002-00276) and the Finnish Graduate School in Electronics Telecommunications and Automation (GETA).

Akeroyd, M. A. (**2001**). ''A binaural cross-correlogram toolbox for MATLAB,'' http://www.biols.susx.ac.uk/home/Michael_Akeroyd/download2.html

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). The CIPIC HRTF Database, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, pp. 99–102.

Bernstein, L. R., and Trahiotis, C. (**1996**). ''The normalized correlation: Accounting for binaural detection across center frequency,'' J. Acoust. Soc. Am. **100**, 3774–3784.

Bernstein, L. R., van de Par, S., and Trahiotis, C. (**1999**). ''The normalized interaural correlation: Accounting for NoSπ thresholds obtained with Gaussian and ''low-noise'' masking noise,'' J. Acoust. Soc. Am. **106**, 870–876.

Blauert, J. (**1971**). ''Localization and the law of the first wavefront in the median plane,'' Acustica **50**, 466–470.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization*, revised ed. (The MIT Press, Cambridge, MA).

Blauert, J., and Cobben, W. (**1978**). ''Some consideration of binaural cross correlation analysis,'' Acustica **39**, 96–104.

Boehnke, S. E., Hall, S. E., and Marquadt, T. (**2002**). ''Detection of static and dynamic changes in interaural correlation,'' J. Acoust. Soc. Am. **112**, 1617–1626.

Braasch, J. (**2002**). ''Localization in the presence of a distracter and reverberation in the frontal horizontal plane. II. Model algorithms,'' Acust. Acta Acust. **88**, 956–969.

Braasch, J. (**2003**). ''Localization in the presence of a distracter and reverberation in the frontal horizontal plane. III. The role of interaural level differences,'' Acust. Acta Acust. **89**, 674–692.

Braasch, J., and Blauert, J. (**2003**). ''The precedence effect for noise bursts of different bandwidths. II. Comparison of model algorithms,'' Acoust. Sci. Tech. **24**, 293–303.

Braasch, J., and Hartung, K. (**2002**). ''Localization in the presence of a distracter and reverberation in the frontal horizontal plane. I. Psychoacoustical data,'' Acust. Acta Acust. **88**, 942–955.

Braasch, J., Blauert, J., and Djelani, T. (**2003**). ''The precedence effect for noise bursts of different bandwidths. I. Psychoacoustical data,'' Acoust. Sci. Tech. **24**(5), 233–241.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**). ''Binaural processing model based on contralateral inhibition. I. Model structure,'' J. Acoust. Soc. Am. **110**, 1074–1088.

Clifton, R. K. (**1987**). ''Breakdown of echo suppression in the precedence effect,'' J. Acoust. Soc. Am. **82**, 1834–1835.

Culling, J. F., Colburn, H. S., and Spurchise, M. (**2001**). ''Interaural correlation sensitivity,'' J. Acoust. Soc. Am. **110**, 1020–1029.

Djelani, T., and Blauert, J. (**2001**). ''Investigations into the build-up and breakdown of the precedence effect,'' Acust. Acta Acust. **87**, 253–261.

Djelani, T., and Blauert, J. (**2002**). ''Modelling the direction-specific build-up of the precedence effect,'' Forum Acusticum, Sevilla, Spain.

Drullman, R., and Bronkhorst, A. W. (**2000**). ''Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation,'' J. Acoust. Soc. Am. **107**, 2224–2235.

Franssen, N. V. (**1960**). Some considerations on the mechanism of directional hearing, Ph.D. thesis, Technische Hogeschool, Delft, The Netherlands.

Freyman, R. L., Clifton, R. K., and Litovsky, R. Y. (**1991**). ''Dynamic processes in the precedence effect,'' J. Acoust. Soc. Am. **90**, 874–884.

Gabriel, K. J., and Colburn, H. S. (**1981**). ''Interaural correlation discrimination: I. Bandwidth and level dependence,'' J. Acoust. Soc. Am. **69**, 1394–1401.
Gaik, W. (**1993**). ''Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling,'' J. Acoust. Soc. Am. **94**, 98–110.
Giguère, C., and Abel, S. M. (**1993**). ''Sound localization: Effects of reverberation time, speaker array, stimulus frequence, and stimulus rise/decay,'' J. Acoust. Soc. Am. **94**, 769–776.
Good, M. D., and Gilkey, R. H. (**1996**). ''Sound localization in noise: The effect of signal to noise ratio,'' J. Acoust. Soc. Am. **99**, 1108–1117.
Good, M. D., Gilkey, R. H., and Ball, J. M. (**1997**). ''The relation between detection in noise and localization in noise in the free field,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 349–376.
Grantham, D. W. (**1982**). ''Detectability of time-varying interaural correlation in narrow-band noise stimuli,'' J. Acoust. Soc. Am. **72**, 1178–1184.
Hartmann, W. M. (**1983**). ''Localization of sound in rooms,'' J. Acoust. Soc. Am. **74**, 1380–1391.
Hartmann, W. M. (**1997**). ''Listening in a room and the precedence effect,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 349–376.
Hartmann, W. M., and Rakerd, B. (**1989**). ''Localization of sound in rooms, IV: The Franssen effect,'' J. Acoust. Soc. Am. **86**, 1366–1373.
Hartung, K., and Trahiotis, C. (**2001**). ''Peripheral auditory processing and investigations of the ''precedence effect'' which utilize successive transient stimuli,'' J. Acoust. Soc. Am. **110**, 1505–1513.
Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). ''Speech intelligibility and localization in a multi-source environment,'' J. Acoust. Soc. Am. **105**, 3436–3448.
IEEE (**1969**). ''IEEE recommended practice for speech quality measurements,'' IEEE Trans. Audio Electroacoust. **17**, 137–148.
ISO 389 (**1975**). ''Acoustics—standard reference zero for the calibration of pure-tone audiometers.''
Jain, M., Gallagher, D. T., Koehnke, J., and Colburn, H. S. (**1991**). ''Fringed correlation discrimination and binaural detection,'' J. Acoust. Soc. Am. **90**, 1918–1926.
Jeffress, L. A. (**1948**). ''A place theory of sound localization,'' J. Comp. Physiol. Psychol. **61**, 468–486.
Koehnke, J., Colburn, H. S., and Durlach, N. I. (**1986**). ''Performance in several binaural-interaction experiments,'' J. Acoust. Soc. Am. **79**, 1558–1562.
Kollmeier, B., and Gilkey, R. H. (**1990**). ''Binaural forward and backward masking: Evidence for sluggishness in binaural detection,'' J. Acoust. Soc. Am. **87**, 1709–1719.
Langendijk, E. H. A., Kistler, D. J., and Wightman, F. L. (**2001**). ''Sound localization in the presence of one or two distracters,'' J. Acoust. Soc. Am. **109**, 2123–2134.
Lindemann, W. (**1986a**). ''Extension of a binaural cross-correlation model by means of contralateral inhibition. I. Simulation of lateralization of stationary signals,'' J. Acoust. Soc. Am. **80**, 1608–1622.
Lindemann, W. (**1986b**). ''Extension of a binaural cross-correlation model by means of contralateral inhibition. II. The law of the first wave front,'' J. Acoust. Soc. Am. **80**, 1623–1630.
Litovsky, R. Y., and Shinn-Cunningham, B. G. (**2001**). ''Investigation of the relationship among three common measures of precedence: Fusion, localization dominance, and discrimination suppression,'' J. Acoust. Soc. Am. **109**, 346–358.
Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). ''The precedence effect,'' J. Acoust. Soc. Am. **106**, 1633–1654.
Litovsky, R. Y., Rakerd, B., Yin, T. C. T., and Hartmann, W. M. (**1997**). ''Psychophysical and physiological evidence for a precedence effect in the median sagittal plane,'' J. Neurophysiol. **77**, 2223–2226.
Lorenzi, C., Gatehouse, S., and Lever, C. (**1999**). ''Sound localization in noise in normal-hearing listeners,'' J. Acoust. Soc. Am. **105**, 1810–1820.
Martin, K. D. (**1997**). ''Echo suppression in a computational model of the precedence effect,'' IEEE ASSP Workshop on Applications of Signal Processes to Audio and Acoustics, New Paltz, NY.
Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture and software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.
Pollack, I., and Trittipoe, W. (**1959a**). ''Binaural listening and interaural noise cross correlation,'' J. Acoust. Soc. Am. **31**, 1250–1252.
Pollack, I., and Trittipoe, W. (**1959b**). ''Interaural noise correlation: Examination of variables,'' J. Acoust. Soc. Am. **31**, 1616–1618.
Rakerd, B., and Hartmann, W. M. (**1985**). ''Localization of sound in rooms. II. The effects of a single reflecting surface,'' J. Acoust. Soc. Am. **78**, 524–533.
Rakerd, B., and Hartmann, W. M. (**1986**). ''Localization of sound in rooms. III. Onset and duration effects,'' J. Acoust. Soc. Am. **80**, 1695–1706.
Rayleigh, L. (**1907**). ''On our perception of sound direction,'' Philos. Mag. **13**, 214–232.
Slaney, M. (**1998**). ''Auditory toolbox: Version 2,'' Technical Report No. 1998-010, http://rvl4.ecn.purdue.edu/~malcolm/interval/1998-010/
van de Par, S., Trahiotis, C., and Bernstein, L. R. (**2001**). ''A consideration of the normalization that is typically included in correlation-based models of binaural detection,'' J. Acoust. Soc. Am. **109**, 830–833.
Viemeister, N. F., and Wakefield, G. H. (**1991**). ''Temporal integragion and multiple looks,'' J. Acoust. Soc. Am. **90**, 858–865.
Wightman, F. L., and Kistler, D. J. (**1992**). ''The dominant role of low-frequency interaural time differences in sound localization,'' J. Acoust. Soc. Am. **91**, 1648–1661.
Yin, T. C. T., and Chan, J. C. K. (**1990**). ''Interaural time sensitivity in medial superior olive of cat,'' J. Neurophysiol. **64**, 465–488.
Zurek, P. M. (**1987**). ''The precedence effect,'' in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York), pp. 85–105.