Failure of the precedence effect with a noise-band vocoder

Bernhard U. Seeber[a] and Ervin R. Hafter

Auditory Perception Laboratory, Department of Psychology, University of California at Berkeley, Berkeley, California 94720-1650

(Received 9 July 2009; revised 9 November 2010; accepted 23 November 2010)

The precedence effect (PE) describes the ability to localize a direct, leading sound correctly when its delayed copy (lag) is present, though not separately audible. The relative contribution of binaural cues in the temporal fine structure (TFS) of lead–lag signals was compared to that of interaural level differences (ILDs) and interaural time differences (ITDs) carried in the envelope. In a localization dominance paradigm participants indicated the spatial location of lead–lag stimuli processed with a binaural noise-band vocoder whose noise carriers introduced random TFS. The PE appeared for noise bursts of 10 ms duration, indicating dominance of envelope information. However, for three test words the PE often failed even at short lead–lag delays, producing two images, one toward the lead and one toward the lag. When interaural correlation in the carrier was increased, the images appeared more centered, but often remained split. Although previous studies suggest dominance of TFS cues, no image is lateralized in accord with the ITD in the TFS. An interpretation in the context of auditory scene analysis is proposed: By replacing the TFS with that of noise the auditory system loses the ability to fuse lead and lag into one object, and thus to show the PE.

© *2011 Acoustical Society of America*. [DOI: 10.1121/1.3531836]

PACS number(s): 43.66.Pn, 43.66.Qp, 43.66.Rq, 43.66.Ts [RYL] Pages: 1509–1521

I. INTRODUCTION

Cochlear implants (CIs) have helped many patients to re-gain the ability to understand speech. However, speech understanding in background noise or reverberation presents a major challenge for individuals with CIs, whose speech reception thresholds are often 10 dB and sometimes even as much as 24 dB higher than those of normal-hearing listeners (e.g., Schön *et al.*, 2002; Cullington and Zeng, 2008). This inability to cope with background sounds is, among other factors, linked to misrepresentation of information in the temporal fine structure (TFS). TFS has recently been connected with the ability to listen into the dips of a modulated masker (Lorenzi *et al.*, 2006; Hopkins and Moore, 2009). Most current CIs stimulate with constant rate electric pulse trains modulated with the stimulus envelope, which does not encode the stimulus TFS (Wilson and Dorman, 2008).

Listeners with bilateral CIs generally show strongly reduced ability for localizing horizontal sound sources, although a few CI listeners have been shown to localize nearly as well in quiet as listeners with normal hearing (Seeber *et al.*, 2004; Litovsky *et al.*, 2009). Localization of two bilateral CI listeners with excellent acuity was recently shown to be based on interaural level differences (ILDs), with interaural time differences (ITDs) carried in the envelope being largely ignored (Seeber and Fastl, 2008). Nevertheless, one might expect performance based solely on ILDs to break down when other sounds are present, as changes in the envelope affect modulation depth and ILDs in such situations (Shinn-Cunningham *et al.*, 2005).

For listeners with normal hearing, ITDs derived from the TFS are the most effective cues for localizing broadband sounds (Macpherson and Middlebrooks, 2002). However, in the presence of reflective surfaces, those cues may be changed during the ongoing stimulus by the addition of echoes. When the direct sound (lead) is followed after a brief delay by its reflection (lag), the percept is of a single source coming from the direction of the first wavefront (Blauert, 1997; Litovsky *et al.*, 1999). Because the lead takes precedence over the direction and audibility of the lag, this has been called the precedence effect (PE) (Wallach *et al.*, 1949). The PE is described by two phenomena which normally occur simultaneously: Fusion of lead and lag into a single object occurs at delays shorter than the echo threshold (ET). Localization dominance relates to the directional aspect of the PE and is seen as the localization of the single, fused image at or toward the lead location. By this definition, fusion is a prerequisite for localization dominance. Auditory scene analysis takes a more general view of fusion and defines it as the assignment of acoustic elements to auditory objects. Here, these are the acoustic elements of lead and lag, which may be fused into a single image.

Previous work looking at the relative contributions of ITDs in the TFS and envelope for localization has found that onset information dominates the PE for short sounds, but, for longer, sufficiently broadband sounds, dominance shifts to the fine structure (Tobias and Schubert, 1959; Rakerd and Hartmann, 1986; Freyman and Zurek, 1997; Hartung and Trahiotis, 2001). Using a chimaerizer, Zeng *et al.* (2004) showed that lateralization of speech sounds can nevertheless be dominated by envelope-ILDs while TFS-ITDs contributed only slightly to the position percept.

Much of our knowledge about the contribution of envelope information to understanding speech stems from the use of a vocoder (Dudley, 1939; Shannon *et al.* 1995). The

[a] Author to whom correspondence should be addressed. Also at: MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD, United Kingdom. Electronic mail: seeber@ihr.mrc.ac.uk

vocoder splits the sound signal into frequency bands, extracts the envelope in each band, and uses these extracted envelopes to modulate narrow-band carriers (e.g., noises) which are then summed and presented to the subject. The processing is similar to that in a CI where the envelopes of each channel would be used to amplitude-modulate electric pulse trains on different electrodes. Electrode location is simulated in the vocoder by the center frequency of the narrow-band carrier, because it corresponds to the region with the highest excitation. Vocoders have been successfully used to study the theoretical number and placement of channels in the implant for optimization of speech understanding in noise (Qin and Oxenham, 2003; Shannon *et al.*, 2004; Fu and Nogaki, 2005; Garadat *et al.*, 2009).

While a monaural vocoder provides a valuable tool for study of perception based on envelopes, one must keep in mind that normal-hearing listeners processing speech in a noisy acoustic background also rely on cues in the TFS for segregating sources on the basis of spatial separation (Nábêlek and Robinson, 1982; Bronkhorst and Plomp, 1988; Edmonds and Culling, 2005). This raises a potential confound when evaluating the usage of binaural vocoders for simulation of bilaterally implanted CIs, because the subject with normal hearing may respond to unavoidable ITDs from the phase of the TFS in the carrier. For example, use of a correlated noise carrier in matched bands in the bilaterally vocoded stimuli produces an ITD in the TFS of zero, a value normally associated with sources on the auditory midline. Conversely, uncorrelated carriers in the matched filters generate ITDs in the TFS indicative of a broad, somewhat amorphous sound (Blauert and Lindemann, 1986).

The present study examined the PE in normal-hearing participants who listened through binaural vocoders. The vocoder eliminated the TFS of the source while keeping its envelope largely intact. The primary focus was on whether ITDs and ILDs in the envelope would provide a PE similar to that found when natural TFS is present. Secondary interest was placed on the importance of binaural phase information based not on the stimulus but rather on the interaural correlation of noise carriers in the vocoders. While the primary motivation was to study the relative importance of binaural cues for creating the PE in conditions related to current CIs, the results could have implications for potential inclusion of TFS in future implants. Rather than relying entirely on discrimination as the measure of performance, subjects described the locations of spatial stimuli with separate responses for the perceived direction of both lead and lag as well as the extent of their fusion into a single auditory object. The PE was examined with unprocessed and vocoded versions of three kinds of stimuli: a low-pass noise burst for which the PE should be based on envelope-ITDs as ILDs are small at low frequencies; a wide-band noise burst, additionally providing ILD cues; and three different speech tokens, each with shifting spectro-temporal structure.

II. METHODS

A. Subjects and stimuli

The same six paid subjects (age: 19–29 yr, median: 20 yr) participated in all experiments but only three of them completed experiment 2 with the words "wide" and "teak." All had normal audiometric thresholds <20 dB hearing level (HL) as assessed with a Bèkèsy tracking procedure within 300 Hz to 10 kHz.

Stimuli were generated prior to the experiments and the same stimulus samples were used across repetitions in all experiments as well as in lead and lag: a burst of white noise (10 ms duration, 1 ms Gaussian rise/decay times, 300 Hz to 10 kHz); a low-pass noise burst (10 ms duration, 1 ms Gaussian rise/decay times, 300–770 Hz); the consonant–vowel–consonant (CVC) words "shape," "wide," and "teak" spoken by a female speaker and taken from the CASPA 3.0 speech test. The durations of the vowels, computed at −60 dB from the rms-maximum, were 940 ms for "shape," 927 ms for "wide," and 751 ms for "teak." Onset times, computed from the rms-level (5 ms exponential low-pass filter) between −40 dB and −10 dB from the stimulus maximum, were 89 ms for "shape," 60 ms for "wide," and 7.3 ms for "teak." The onset time for "teak" depends highly on the definition of the threshold because the plosive [t] produces a sharp first maximum 7 dB below the later occurring maximum of the vowel. In each trial the level of the stimulus was roved in 2 dB steps within ±6 dB from a base level of 60 dB(A) for noise and 55 dB(A) for the CVCs. Ten trials were collected for each stimulus for the lead to the left and to the right.

B. Binaural stimuli with and without noise-band vocoding

Experiments were done in a virtual listening environment with headphones (Sennheiser HD 580, diffuse field equalized Sennheiser, Wennebostel, Germany) using individually selected head-related transfer-functions (HRTFs). In a two-step procedure participants selected an HRTF-pair from a catalog of non-individual HRTFs by first selecting five HRTFs from the catalog with a general question and then choosing one HRTF from among these five according to multiple criteria. A previous study showed that this 10-min procedure finds an HRTF which generally yields smaller localization variance and increased externalization (Seeber and Fastl, 2003). A preliminary study with the same participants showed only small differences between the PE found with open-field listening and with headphone listening based on selected HRTFs.

Stimuli were first filtered with HRTFs for the left and the right ear to generate virtual stimuli for headphone presentation with the lead from +30° and the lag from −30° azimuth or vice versa. Filtered lead and lag stimuli were then summed separately for each ear to create signals at the two ears similar to those when sounds are played from loudspeakers in the free-field. The filtered stimuli carried natural, frequency-dependent ILDs as well as ITDs in the envelope and the TFS. In experiment 1 these signals were played directly over headphones.

In experiments 2 and 3, signals were next passed through a binaural noise-band vocoder with 16 channels. Figure 1 illustrates processing in the right channel of the vocoder (the left channel was identical). Signals for each ear were bandpass filtered (6th-order Chebyshev) in 16 logarithmically spaced channels from 300 Hz to 8 kHz ("analysis filters") giving each a bandwidth approximately equal to the critical bandwidth of the corresponding auditory filter

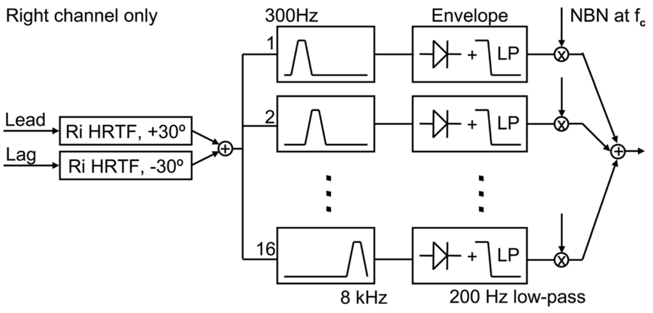

FIG. 1. Schematic of the noise-band vocoder used to process the lead–lag stimuli which were spatialized with HRTFs. Interaural correlation of the carrier noise was varied (cf. Sec. II D; LP, low-pass filter; NBN, narrow-band noise).

(Zwicker and Fastl, 1999). The envelope was computed from the signal in each channel by full-wave rectification and low-pass filtering at 200 Hz (8th-order Butterworth). The envelope of each channel was multiplied with a narrow band of noise ("carrier") with the same center frequency and bandwidth as the analysis filter of that channel. The 16 modulated noise bands for each side were then summed to create signals to be played to the right and left ears via headphones. Processing was digital at a sampling rate of 44.1 kHz. Because HRTFs were individually selected by the subjects, vocoded stimuli were generated for each subject with new noise carriers prior to the experiments and stored to disk. The same stimulus token was then used across trials.

As discussed in the Introduction, normal-hearing listeners are highly sensitive to the interaural correlation of the carrier noise. Uncorrelated noise is perceived with a diffuse location inside the head with a tendency toward hearing two auditory images, one at each ear, while correlated noise gives rise to a compact auditory image in the center of the head (Blauert and Lindemann, 1986). In experiments 2 and 3, the envelopes extracted independently for the vocoder at each ear were applied to noise-band carriers with interaural correlations of 0.00, 0.35, 0.70, 0.90, or 1.00. Note that correlations in the TFS of the vocoded stimuli were not related to the original stimuli, either noises or words, but represent interaural phase which does not affect perception of stimuli in either channel alone. Note also that the channel envelopes extracted in the vocoder contained ILDs which, across channels, followed their natural, frequency-specific course. ILDs were reproduced with little error because carriers had the same level on both ears. Since there was no temporal quantization of the channel envelopes, envelope ITDs, such as those in onsets, were likewise well reproduced.

C. Localization tests

A visual display was projected in front of the subject on an acoustically transparent curtain (Seeber *et al.*, 2010). In each trial, 0.5 s after the auditory stimulus, a visual marker was presented directly in front of the listener at 0°. The subject then used a trackball to move the marker horizontally to a place indicating the perceived azimuth or the lateralized position before confirming the response with a button press (Seeber, 2002). The total allowable range in azimuth was ±37°. Sounds not processed through vocoders (experiment 1) were expected to be localized outside the head by virtue of the HRTF processing and the marker was the two-dimensional projection of a three-dimensional red ball presented on an entirely black background. Sounds presented through the vocoder (experiments 2 and 3) were expected to be heard more "inside of the head." For these sounds, the visual display showed a horizontal white line, terminated by vertical strips labeled "left ear" and "right ear" to which the numbers −1 and +1 were assigned. Participants marked the lateralized position with a red vertical line-segment.

D. PE test procedures

The PE was investigated in a localization dominance experiment in which lead and lag were played from opposite sides at ±30° virtual azimuth with a probability of 0.5 that the lead would be on the left. In each case, the lag was a delayed copy of the lead. In a single experimental session, subjects were told that if they heard only one image, they were to indicate its azimuth. However, if they heard two images, they should mark the leftmost image (Hafter and Jeffress, 1968; Litovsky and Shinn-Cunningham, 2001). Randomizing the side of lead and lag sounds on every trial meant that subjects responded to the lead on one half of the trials and the lag on the other half. As a precaution against biases, in mirror conditions subjects were instructed to point to the rightmost image. The two sets were combined for plotting by changing the signs of responses for trials in which the instruction was to point to the leftmost image and the lead was on the left and combining these flipped data with those when pointing was to the rightmost image and the lead was on the right, arbitrarily labeling the lead location as +30° and the lag location −30° azimuth. Likewise, the lag image was extracted by combining the results for pointing to the leftmost image when the lag was on the left with sign-inverted data for pointing to the rightmost image when the lag was on the right. The results of experiments 2 and 3 using the lateralization procedure were combined such that responses to the lead were labeled toward +1 and responses to the lag toward −1. In experiment 3 subjects were asked to localize either the most dominant or the weakest image instead of the leftmost or rightmost.

When confirming the localized position of the sound by pressing a button on the trackball, subjects were instructed to choose the left button if they perceived one sound image and the right button if two or more images were perceived. These results indicate a *subjective* impression of whether the lag was audible. Data were analyzed separately for the effective instruction to point to the lead or the lag image. The results on perceived fusion are plotted in the lower part of most figures.

In all experiments the same lead–lag delays were tested. These were 0, 0.5, 2, 4, 7, 11, and 16 ms for the wide-band noise burst, 0, 0.5, 2, 6, 12, 20, and 30 ms for the low-pass noise burst and 0, 0.5, 2, 6, 12, 24, and 48 ms for the three words.

III. EXPERIMENT 1: PE IN THE VIRTUAL FREE-FIELD

A. Overview and procedures

The two noise stimuli and the word "shape" were used to measure the PE in a virtual free-field produced with headphones and HRTFs. These data would form the baseline against which to interpret the results with vocoders in experiments 2 and 3. Ten localization responses were collected at each of the delays listed in Sec. II C for each lead direction and for each stimulus, giving a total of 420 trials (10 × 7 × 2 × 3) per instruction (point to the rightmost or leftmost image). Sound presentation was randomized across stimuli, lead directions, delays, and repetitions. The experiment was broken into runs of circa 110 trials lasting about 8 min. The first three trials in each run were disregarded. Subjects received training of at least 20 min before data collection began.

B. Results

Figure 2 presents pooled results across all subjects for the three stimuli. The top panel shows individual responses and their medians (connected by lines) separately for the lead and lag image. Error bars depict quartiles. The lower panel presents the percentage of trials in which subjects reported hearing more than one image.

Results for the low-pass noise burst (left column of Fig. 2) followed the well-known pattern for the PE: At zero delay and very short delays (0.5 ms), one image was localized near the center (0°) between both (virtual) loudspeakers, a pattern termed “summing localization.” The image shifted close to the lead location at 2 ms with only a few aberrant responses at the location of the lag. This was true even though subjects were asked to point toward a possible image on the lag side. The appearance of only one image localized at the lead is called “localization dominance” or the “precedence effect” (PE). The bias of the image toward the middle indicates a remaining influence of the lag, although less so at 2 ms than at 0.5 ms. Increasing the lead–lag delay produced more responses at the lag location, although median responses at 6 and 12 ms were near the lead, even with instructions to point to the lag. Starting at 20 ms delay, median responses separated into one image at the lead and one at the lag location, indicating that the lag was now separately audible as an echo (Blauert, 1997; Litovsky *et al.*, 1999).

The lower panels of Fig. 2 give the percentage of trials in which subjects reported hearing two or more images. For the low-pass noise burst, this percentage increased monotonically with increasing delay, consistent with results from the localization test. The 50% mark fell between 6 and 12 ms delay. The fact that the slope of this function was shallow suggests that the “ET,” the transition between hearing one image or two, was not well defined for the low-pass noise or was highly variable across subjects.

The results for the wide-band noise burst generally resembled those for the low-pass noise burst (middle column of Fig. 2). One difference is that the lag affected localization responses at shorter delays: The median of the responses to the lag crossed the midline (0°) at 7 ms, while for the low-pass noise it moved to the lag side between 12 and 20 ms delay.

Results for the word “shape” also exhibited a region (0.5–12 ms) of strong lead dominance with a defined crossover point of 50% for the lag between 12 and 24 ms. Similarly, the crossover from hearing one image to hearing two or more images was well defined at about the same delay (lower panel).

In summary, the PE was active for all three unprocessed stimuli and could be seen at short delays as an almost exclusive localization at the lead.

IV. EXPERIMENT 2: PE WITH A NOISE-BAND VOCODER

A. Overview and procedures

The procedures here were essentially the same as those in the virtual free-field test of experiment 1 with the primary difference being the processing of stimuli though the binaural vocoders described in Sec. II D. Again, ten localization responses were collected for each delay, lead direction and

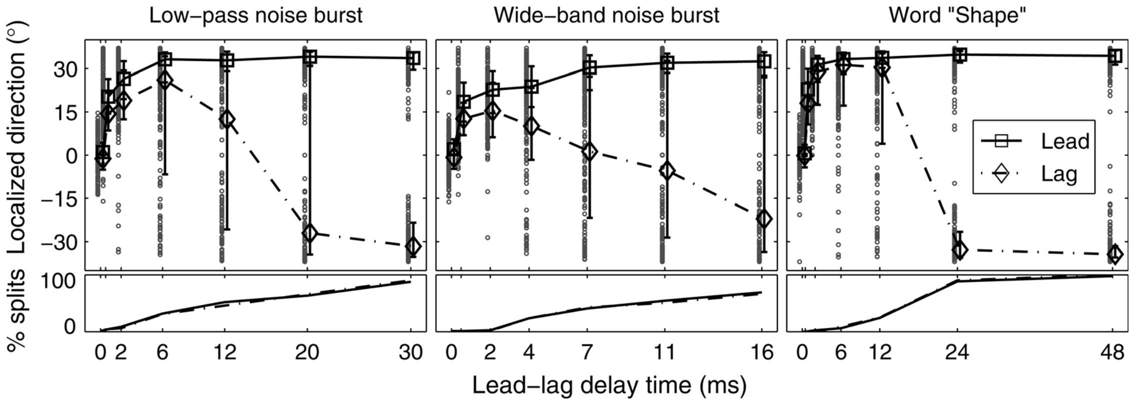

FIG. 2. Top: Localization results as a function of the delay between lead and lag for a low-pass noise burst (left column), a wide-band noise burst (middle), and the word “shape” (right). As plotted, the lead was played from +30° and the lag from −30° azimuth, but in the experiment lead and lag locations were randomized. In one session subjects were instructed to point to the rightmost image if they heard two or more sounds, and in another session to the leftmost image. When the lead was on the left and subjects were instructed to point to the leftmost image, they effectively pointed to the lead. These data were sign-inverted and combined with the data for pointing to the right when the lead was on the right and plotted as the lead image. Likewise, the lag image was formed from data when the instruction was to point to the leftmost image and the lag was on the left. 120 responses from six subjects are plotted per delay. Medians and quartiles of the pooled results are given for the lead and lag image. Bottom: Subjects were instructed to press a different button according to whether they perceived one or more than one image, i.e. whether the lag was audible. The lower subplots show the percentage of responses for hearing two or more images. The solid line indicates data from pointing to the lead while the dashed-dotted line indicates those from the lag. Only one line is visible because of their coincidence.

stimulus and additionally for each of the five interaural correlations (0.00, 0.35, 0.70, 0.90, and 1.00) of the carrier noises (10 × 7 × 2 × 3 × 5 = 2100 trials per instruction). All trials were presented in randomized order across 17 runs of ca. 126 trials each. As noted in Sec. II, preliminary testing showed that vocoded sounds were perceived "inside of the head." Subjects indicated the lateralized position using the visual display which was terminated with the labels "left ear" and "right ear," represented by the numbers −1 and +1, respectively. Subjects participated in experiment 1 before starting experiment 2 to ensure sufficient training in the task.

When the PE failed for the word "shape," additional testing was done with two other CVCs: The word "wide" as an example of a stimulus with a long, steady vowel, and the word "teak" with its plosive onset. The number of trials and the vocoder processing were the same but only four values of interaural carrier correlation were tested (0.0, 0.4, 0.8, and 1.0).

B. Results with low-pass and wide-band noise bursts

Figure 3 presents results for the vocoded low-pass and wide-band noise bursts (left and right column, respectively) with uncorrelated (top) and correlated carriers (bottom). Positive values indicate a response toward the ear at the side of the lead and negative responses toward the ear at the side of the lag. The lower panels show the percentage of trials in which listeners reported hearing two or more images. As before, lateralization results are presented as medians and quartiles of pooled responses for pointing to the lead (□) or the lag (◇) image (cf. Sec. II C). The results for intermediate correlations appeared as if they were scaled versions of those for correlations zero and one and were thus omitted for brevity in this and other plots.

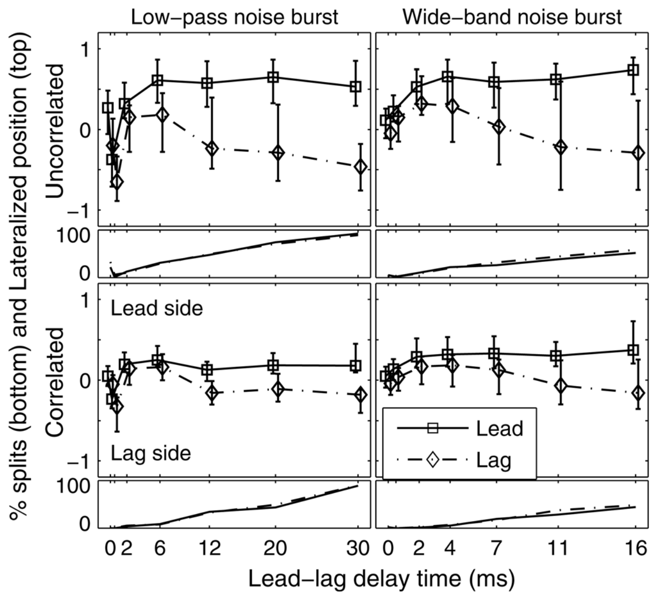

FIG. 3. Lateralization results for vocoded lead–lag stimuli; a low-pass noise burst (left column) and a wide-band noise burst (right). Results are plotted similarly to Fig. 2 as medians and quartiles of responses parsed for pointing to the lead or lag image. Note that ±1 depicts maximum lateralization at the ears and not the lead or lag position. The top row presents results for uncorrelated carriers and the bottom row for correlated carriers. The lower subplots show the percentage of responses for hearing two or more images with the solid line indicating data from pointing to the lead and the dashed-dotted line those from the lag. Note that both lines are almost coincident.

Results for the low-pass noise burst showed the PE, i.e., lateralization toward the lead, at delays of 2–6 ms. Note that the lead carried binaural cues relating to +30° which would not result in complete lateralization at the right ear (+1) but at a somewhat more central location (Blauert, 1997). Especially interesting is that despite the diffuse spatial appearance expected for the uncorrelated noise carriers (Blauert and Lindemann, 1986), lateralization of uncorrelated carriers with vocoder-envelopes taken from the original stimuli yielded relatively compact responses similar to those in the simulated free-field (experiment 1), with only few responses at the lag for delays between 2 and 6 ms. Subjects responded as hearing more than one image in 11% of the trials at 2 ms delay—only slightly more than the 7% without vocoder.

Responses toward lead and lag were closer together and more centered with correlated carriers, even at long delays. Because unmodulated correlated carriers should give rise to a focused sound image heard at the center of the head (ITD = 0) (Blauert and Lindemann, 1986), this bias toward the center seems to reflect the influence of the correlated carrier. This bias would affect the lead when presented alone and it can be expected that it would have a similar effect when lead and lag are presented together. Nevertheless, despite the binaural effect of the carriers' phase, localization dominance was still evident as the medians of responses to lead and lag coincided at a position on the lead side for delays between 2 and 6 ms, suggesting that binaural information of the lead carried in the envelope was accessed to position the fused image. At the 2 ms delay subjects reported hearing two or more images on only 4% of the trials, slightly less frequent than for unprocessed stimuli (7%).

Responses for 0.5 ms delay were consistently toward the location of the lag irrespective of carrier correlation. This "reversed" PE has been previously studied and termed "anomalous localization." It occurs for narrow-band lead–lag stimuli when phase cancellation results in a substantial ILD favoring the lag, here of 8 dB (Blauert and Cobben, 1978; Tollin and Henning, 1999). The vocoder passes on the "erroneous" ILD and in the absence of stimulus-related TFS the auditory system lacks the information to resolve the anomaly.

Results with the wide-band noise burst were qualitatively similar to those with the low-pass noise with the exception that anomalous localization did not occur (Fig. 3, right, note the different time axis). Although the results for the correlated carrier showed far less lateralization than those for the uncorrelated carrier, the pattern of responses was similar. Median responses for lead and lag images were on the side of the lead for 2–4 ms delay regardless of carrier correlation. Subjects also rarely reported hearing two or more images at those delays. The qualitative similarity to the results with unprocessed stimuli in Fig. 2 shows that the PE functioned, despite vocoding.

In summary, localization (lateralization) dominance was observed for brief noise stimuli processed with a noise-band vocoder, although anomalous lateralization, where the lag dominates, occurred for low-pass noise at very short

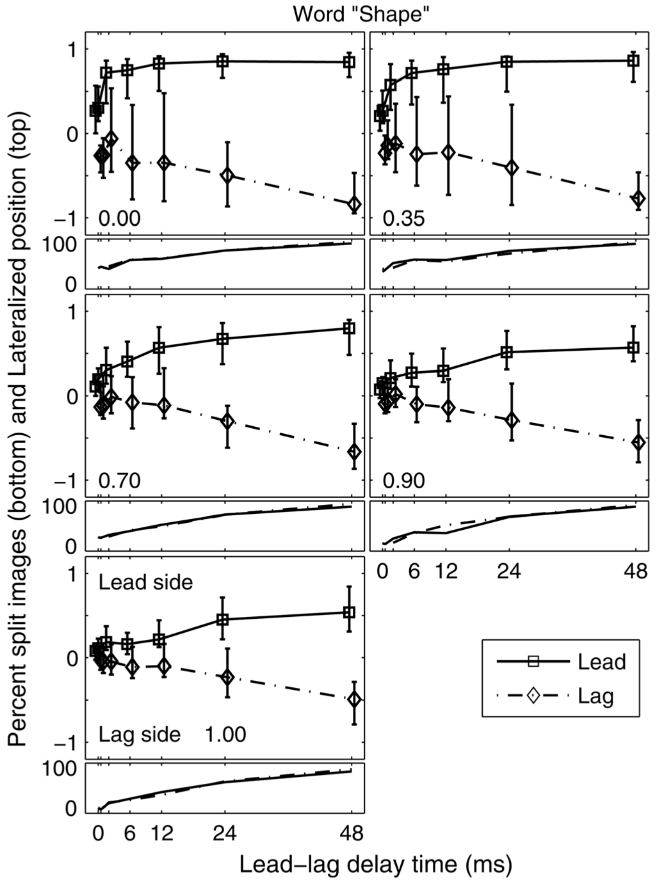

FIG. 4. Same as Fig. 3, but for the word "shape" and five different values of the carrier correlation given in the inset. Irrespective of carrier correlation, a lag image was reported at almost all delays, indicating that the PE failed.

lead–lag delays. The use of interaurally correlated noise carriers in the binaural vocoder affected the perceived location of the lead and lag images, moving them closer to the midline for higher correlations.

C. Results with the word "shape"

Results for the word "shape" are presented in Fig. 4 for all tested carrier correlations. They differ strongly from results with noise bursts. Irrespective of carrier correlation, lateralization results showed two separate images, one toward the lead and one toward the lag. This suggests that a second image was audible, indicating that lead and lag were not fused as they would be under conditions of the PE. Comparison to Fig. 2 shows that while there were very few reports of hearing more than one image for delays up to 12 ms when the word was not vocoded (0% at 0.5 and 2 ms delay, 8% at 6 ms), there was a substantial number of reported split images after vocoding (uncorrelated carrier: 45% at 0.5 ms, 43% at 2 ms, and 58% at 6 ms delay). This held true for vocoding with correlated carriers (6% at 0.5 ms, 20% at 2 ms, and 27% at 6 ms), which is even more surprising given that the correlated carriers would, by themselves, evoke a single fused image. This suggests that in many trials, at short lead–lag delays, subjects not only indicated two lateralized images, but perceptually segregated lead and lag into two separate auditory objects. Because the PE failed for the word "shape" when processed with a vocoder, we attempted to reproduce the effect using other words.

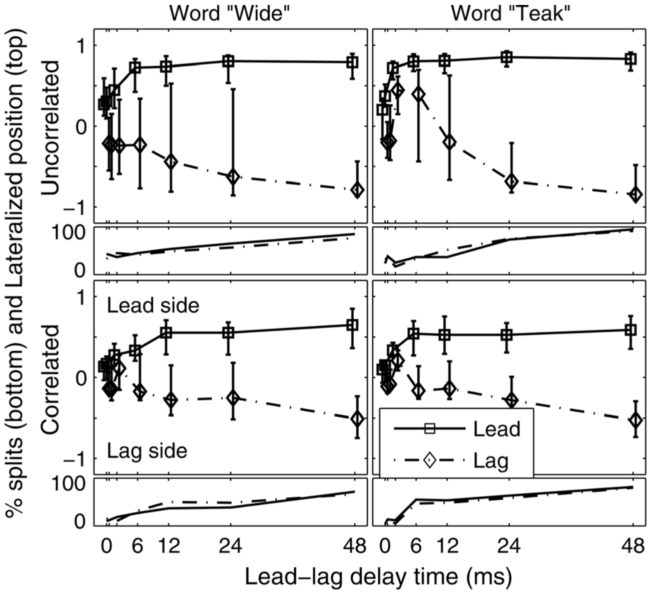

FIG. 5. Same as Fig. 4, but for the words "wide" (left) and "teak" (right) for uncorrelated (top) and correlated carriers (bottom). The lag appeared clearly separated from 2 to 6 ms delay for "wide" and from 6 to 12 ms for "teak."

D. Results with the words "wide" and "teak"

Figure 5 presents results with the words "wide" (left) and "teak" (right) vocoded with an interaural correlation in the carrier of 0.0 (top) and 1.0 (bottom). Results for intermediate correlations and for unprocessed conditions were similar to those for the word "shape" and so are not presented here. The similarity in responses to the three words is also evident for the two extreme correlations (Fig. 5 vs Fig. 4). Subjects lateralized two distinct images at both the lead and the lag starting at 2–6 ms delay for the word "wide" and at 6–12 ms for the word "teak." Particularly with an uncorrelated carrier there was evidence for the PE at 2 ms delay for the word "teak," potentially because of the sharp onset (Miller *et al.*, 2009), although it began to fail at delays as short as 6 ms. The presence of a second image at the side of the lag at these short delays is unlike the PE seen without vocoding which was still stable at 12 ms delay. Subjective reports of hearing two or more images at delays of 6 and 12 ms confirmed the failure of the PE after vocoding (unprocessed/vocoded-correlated/vocoded-uncorrelated carriers, "wide": 0/29/43% at 6 ms, 7/43/51% at 12 ms; "teak": 1/51/35% at 6 ms, and 12/51/44% at 12 ms).

The results with speech tokens demonstrated that the PE did not operate in a stable manner for words processed with a noise-band vocoder.

V. EXPERIMENT 3: THE PE DOES NOT FAIL COMPLETELY

A. Overview and procedures

Experiment 2 showed that two images, one on the lead side and one on the lag side, were usually reported for the

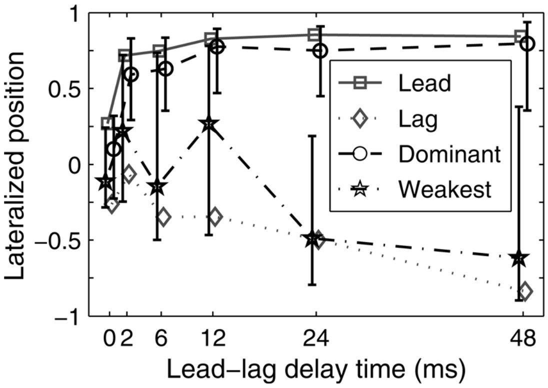

FIG. 6. Lateralization dominance results for the word "shape" for uncorrelated carriers when subjects were instructed to point to the most dominant and the weakest image. Results for the lead and lag image were replotted from Fig. 4 in gray without error bars. The lead and dominant images coincided while the weakest image largely matched the lag image, indicating reduced perceptual salience of the lag.

vocoded words. Our informal observation was that the two images did not always appear to have the same salience and loudness. Rather, the lag image seemed softer and was perceived as being weaker than the lead image. This would suggest that although lead and lag were not fused into one object, the lead nevertheless exerted some "dominance" over the lag. Experiment 3 was designed to investigate the relative salience of the two images. Tests were identical to those in experiment 2 except that subjects were instructed to point either to the "dominant sound image" or the "weakest image" instead of the left/rightmost image.

B. Results

Figure 6 presents the results for the word "shape" with uncorrelated noise carriers. Also shown are results from pointing to the lead and lag image for comparison, replotted from Fig. 4 in gray and without error bars. Median responses to the dominant sound image were nearly identical to the responses to the lead image, suggesting that the lead was perceived as being more salient than the lag. Conversely, in the majority of cases, the weakest image coincided with the lag image. However, error bars of both images overlapped, indicating that the difference in their salience may be small.

Results for the lead, the lag, the dominant, and the weakest image were compared with an analysis of variance (ANOVA) with the factors lead–lag (lead vs lag image location), instruction (lead vs dominant and lag vs weakest), delay (7), and correlation (5). All main factors were significant, but the factor instruction had the smallest effect size and significance [$F(1{,}13999) = 10.36$, $p < 0.0014$, $\eta^2 = 0.00049$, cf. factor lead–lag: $F(1{,}13999) = 3862$, $p < 0.0001$, $\eta^2 = 0.182$]. Of the two-way interactions with instruction, only that with lead–lag reached significance [$F(1{,}13999) = 488$, $p < 0.0001$, $\eta^2 = 0.023$]. Although there was a significant effect of instruction, it's relatively weak significance and effect size suggests that the distributions for the lead and the dominant image as well as for the lag and the weakest image largely overlapped.

In summary, experiment 3 showed that subjects tended to perceive the lead image as being more dominant than the lag image, suggesting that although lead and lag were not fused, information from the lead was accessed to reduce the salience of the lag. This bears similarity with the PE of unprocessed stimuli where the lag appears as a faint echo at delays just above the ET. The results of the present study could thus be interpreted as a drastic shortening of ETs to 2–6 ms for vocoded speech stimuli. In Sec. VI the results for all test stimuli are analyzed to quantify the effects of vocoding on the PE.

VI. ANALYSIS OF THE EFFECT OF VOCODING

In this section results from unprocessed and vocoded conditions are compared to quantify the impact of vocoding. A localization method was used in the tests with unprocessed stimuli, while a lateralization method was used for vocoded stimuli. The methodological differences prevented a direct comparison of localization with lateralization data. The analysis thus focuses on deduced parameters such as the number of responses on the lag side or ETs.

Figure 7 presents the proportion of responses on the side of the lag. A value of 0.5 indicates that there were as many responses on the lead as on the lag side, as would be expected at long delays. When the PE is active, most responses should be on the lead side and those falling on the

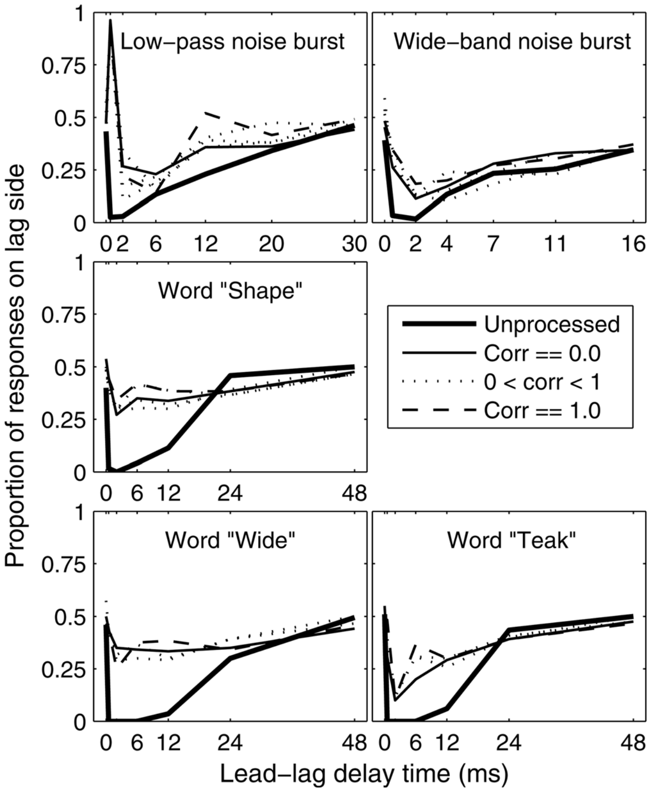

FIG. 7. Proportion of responses on the side of the lag for all stimuli and conditions as a function of delay time. The thick line refers to results with unprocessed stimuli, the thin line to vocoded conditions with uncorrelated carriers, the dashed line with correlated carriers and the dotted lines with carriers with correlations between zero and one. For unprocessed stimuli almost no responses fell on the side of the lag at short delays, indicating strong localization dominance of the lead. After vocoding, many responses occurred on the lag side at similar delays, indicating strongly reduced lead dominance.

lag side should approach zero. This was the case for unprocessed stimuli (thick lines) at lead–lag delays of 0.5, 2, and, to a slightly lesser extent, 6 ms for the noise bursts and for delays up to 12 ms for the words. Vocoding increased the number of responses occurring at the lag side. A total of 15–25% of responses were on the lag side for the noise bursts compared to 30–40% at 6 and 12 ms delay for the words. Thus, at these delays around a third of the responses occurred toward the lag, indicating highly reduced dominance of the lead. For the words "shape" and "wide" already at 2 ms delay a high number of responses were visible on the lag side while for the word "teak" lead dominance was evident at the same delay. Anomalous, almost exclusive localization occurred at the lag for the low-pass noise burst at 0.5 ms delay. There were no clear trends regarding the effect of carrier correlation on the number of responses at the lag side.

ETs computed similarly from localization and lateralization data provide another means by which to compare results across processing conditions. The first measure considered the location of the lag image and reflected the delay at which the lag image crossed the midline from the lag to the lead side [location-based echo threshold (L-ET)]. Third-order polynomials were fitted to the medians of the responses to the lag image and the delays at which they crossed zero, i.e. the midline, were found. L-ETs reflect the smallest delay equal to or greater than 2 ms for which the medians crossed from the lag to the lead side moving from large to short delays.

L-ETs are presented as solid lines in Fig. 8. Horizontal solid lines without markers present unprocessed conditions while L-ETs from vocoded conditions are given as curves with markers as a function of carrier correlation. The L-ET computed for the unprocessed low-pass noise burst was 15 ms and this reduced slightly to about 9 ms after vocoding. For wide-band noise bursts L-ETs increased slightly from 8 to 14 ms after vocoding for medium carrier correlations, but remained roughly unchanged for low and high carrier correlations. The picture was different for all processed words, for which L-ETs were much lower after vocoding. For all unprocessed words L-ETs were 18 ms as the lag image crossed from lead to lag between 12 and 24 ms delay. After vocoding the word "shape," L-ETs could not be computed at any carrier correlation except 0.9 as the lag images did not cross over to the lead side. For the words "wide" and "teak" L-ETs were below 5 ms after vocoding at all but one carrier correlation—much reduced from 18 ms. In summary, the analysis shows that L-ETs were generally little affected by vocoding for the wide-band noise burst, were slightly reduced for the low-pass noise burst, and strongly reduced or even non-measurable for the words.

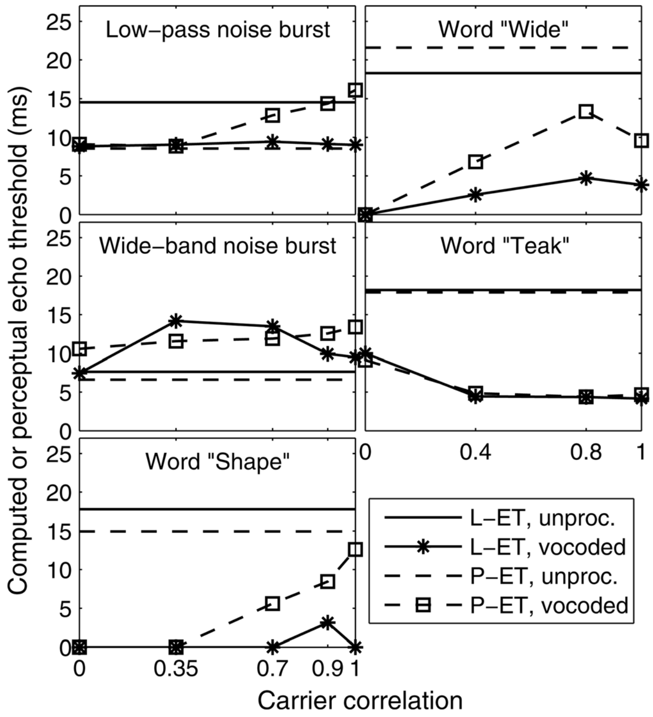

FIG. 8. ETs from unprocessed (horizontal lines without markers) and vocoded (with markers) conditions computed from either localization data (L-ET, solid lines) or from responses for hearing more than one image (P-ETs, dashed lines). For vocoded conditions ETs are given as a function of carrier correlation.

A second measure, perceptual ETs (P-ETs), was derived from the subjective fusion ratings (Fig. 8). P-ETs were defined as the delay at which responses for hearing split images exceeded 40% on average. A threshold of 40% was chosen because ETs could not be determined for all conditions at the 50% level. P-ETs were similar to L-ETs for unprocessed conditions and generally followed similar patterns after vocoding. P-ETs increased slightly after vocoding for both noise bursts, but were lower or even zero for the words. For the low-pass noise and the words "shape" and "wide" P-ETs increased with carrier correlation. The similarity of P-ETs and L-ETs indicates that localization dominance and fusion measures were related.

Cluster analysis was used to ascertain if responses were bimodally distributed or if they stemmed from a single normal distribution. The expectation–maximization algorithm was used to fit one or two normal distributions to the pooled localization results for the lead and the lag image of each subject. Using a loglikelihood-test it was determined if two clusters explained the data significantly better than the fit with one cluster (at $p = 0.01$ for chi^2-distribution with three degrees of freedom). Since this was also often significant for skewed distributions producing nearby clusters, one cluster was required to be on the left hemisphere and the other on the right.

Figure 9 reports the proportion of subjects for which two clusters fit the data significantly better than one cluster, i.e. for which images were split. For the low-pass noise there were generally fewer split images with vocoding than without. The reason for this lies partly in the way split images were computed. Images had to be on opposite sides to be counted as split, but the lag image was on the lead side at 6 ms delay (cf. Fig. 3). However, lateralization variance was also increased for both images, leading to broader distributions which came close together for highly correlated carriers. This casts doubt that at least with high correlations two images were lateralized, in agreement with the few subjective reports for two images perceived. For the wide-band noise burst there were only small differences between unprocessed and vocoded conditions, although there was a tendency toward more split images at 2 and 4 ms delay after

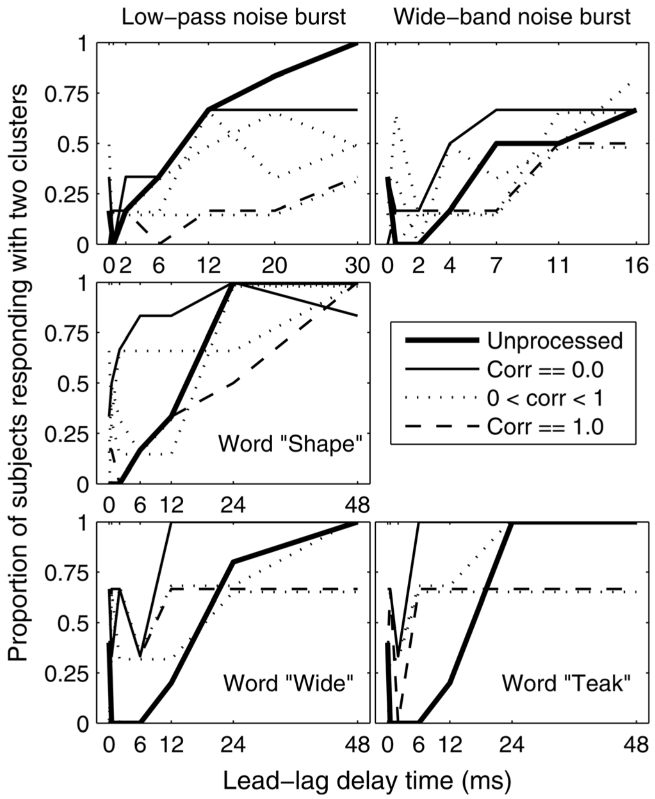

FIG. 9. Proportion of participants for which statistical testing indicated that responses cluster significantly into two images rather than one as a function of delay time. Cluster analysis was used to fit one or two Normal distributions to the pooled results for the lead and lag image of each subject. Given is the proportion of subjects for which the fit with two clusters resulted in a significantly higher loglikelihood than that with one cluster ($p < 0.01$).

vocoding. More split images were reported after vocoding for the word "shape" at short delays, but only for the two lowest carrier correlations (0, 0.3). For higher correlations there was not much difference from the unprocessed case. For the words "wide" and "teak," however, split images occurred more frequently for delays of 2–12 ms after vocoding, regardless of carrier correlation, with "teak" at 2 ms being the only exception. This shows that at delays at which the PE was active for unprocessed stimuli, one to two thirds of the subjects responded with a split distribution of two separate images for vocoded stimuli. In conclusion, subjects frequently failed to show the PE with three vocoded test words at delays of 2–12 ms, quantitatively demonstrated through a high number of responses on the lag side, through reduced computed ETs, and through the frequent occurrence of bimodally distributed localization responses.

A fourth analysis tested if results for vocoded conditions differed between words. An ANOVA with factors word (3) × lead–lag (2) × delay (7) × correlation ($=0|1$) was conducted on the pooled lateralization results of the three subjects who completed the experiment with all words. All main factors were highly significant. Lead–lag explained the largest variance [$F(1,5039) = 3787$, $\eta^2 = 0.35$, $p < 0.001$] while word was the factor with the smallest effect size η [$F(2,5039) = 15.7$, $\eta^2 = 0.0029$, $p < 0.001$]. All two-way interactions were significant ($p < 0.001$) but the interactions including word had the smallest effect sizes [largest was word × delay: $F(12,5039) = 4.34$, $\eta^2 = 0.0049$, $p < 0.001$]. The main factor word became insignificant when the results for the word "teak" were taken out [$F(1,3359) = 0.31$, $\eta^2 = 0.00004$, $p = 0.58$], as did the interaction between word and delay [$F(6,3359) = 2.18$, $\eta^2 = 0.0019$, $p = 0.04$], although interactions with other factors retained significance. Since the factor word remained significant when either of the other words was removed, results for "teak" seemed to differ from those of "shape" and "wide." The factor word also became insignificant when the results for 2 and 6 ms delay were removed [$F(2,3599) = 1.97$, $\eta^2 = 0.00047$, $p = 0.14$]. This is in agreement with the observation that the PE operated at short delays for "teak" despite vocoding, but not for the other words (cf. Figs. 4 and 5). As results did not differ between words except at very short delays and the factor word had the smallest effect size of all factors, it can be concluded that the PE generally failed in a similar way for all chosen words.

VII. OVERALL DISCUSSION

This study investigated the PE in a localization dominance experiment for stimuli presented with and without processing through a noise-band vocoder. A noise-band vocoder replaced the TFS of the signal with the random fine structure of noise while largely maintaining the temporal envelope of the signal. Without vocoder processing, localization dominance was observed at short delay times for all test stimuli, namely a low-pass noise burst, a wide-band noise burst, and a speech token (experiment 1). With processing, localization dominance was still strong for a wide-band noise burst and for a low-pass noise burst, although more responses were given on the lag side. Anomalous localization at the lag occurred for the low-pass noise burst at a lead–lag delay of 2 ms in the vocoded condition only. For tests with speech stimuli, vocoding caused the PE to fail at 6 and 12 ms delay and for some words even at 2 ms. Subjects frequently lateralized two sound images, one at the lead side and one at the lag side (experiment 2). The split occurred regardless of carrier correlation and with test words with different envelope features, suggesting that it occurs for longer duration stimuli or speech in general. Carrier correlation affected the extent of lateralization of the images. The failure of the PE with speech stimuli at short delays was also reflected in lower subjective ETs. When the lag was not completely suppressed it was often perceived as being weaker than the lead (experiment 3).

Traditionally, the PE has been discussed from two viewpoints: (1) binaural cues that give rise to the PE and (2) the contribution of temporal envelope information with a particular focus on onsets. More relevant to CI-listening, we discuss the removal of TFS by the vocoder by asking the following questions: What were the relative contributions to the PE of information in the envelope and the TFS? After vocoding, why did the PE operate more effectively for brief noise bursts than for longer duration speech stimuli? We argue that rather than restrict the discussion to envelope vs TFS, the PE should be considered in the context of auditory scene analysis.

A. Evaluation of binaural cues

A multitude of studies has investigated the contribution of binaural information to the localization of a single sound, leading to the formulation of the duplex theory and its extensions. Based on these, it is thought that ITDs at low frequencies, computed from the fine structure or phase of the signal, dominate localization of broadband and low-frequency sounds (Macpherson and Middlebrooks, 2002). ITDs can be derived from the envelope in high-pass sounds; however, their relative weighting seems to be low (Henning, 1974; Nuetzel and Hafter, 1981). Instead, ILDs are thought to provide localization cues for high-frequency sounds. If one were to assume that this weighting of the cues holds for the PE, ITDs at low frequencies would be crucial. The vocoder used noise carriers which contain ITDs in the TFS. These ITDs were unrelated to the original stimuli, like the carriers themselves, and this might be one reason why the PE failed. On the other hand, information for the PE was carried in envelope ITDs and ILDs and the fact that the PE occurred for brief stimuli or short delay times suggests that this information was, indeed, accessed. Evidence for the dominance of ILDs over ITDs in vocoded stimuli comes from Zeng *et al.* (2004) and it seems that more discussion is needed about cue dominance in conditions with altered binaural cues and when more than one sound (e.g., a lead and a lag) is present.

Discrimination suppression experiments have shown that access to both binaural cues, ITDs and ILDs, is more difficult in the lag when it is preceded by the lead (Zurek, 1980; cf. Houtgast and Aoki, 1994; Litovsky *et al.*, 1999). In the context of listening in rooms, ITDs, particularly when evaluated from the envelope of the broadband signal, carry reliable information (Hartmann, 1983; Shinn-Cunningham *et al.*, 2005). However, there have been only few models that explain the PE on the basis of envelope cues, or that make extended use of them (Zurek, 1987) and there has been only a few studies characterizing the influence of envelope cues beyond the initial onset (Rakerd and Hartmann, 1986; Hafter and Buell, 1990; Freyman and Zurek, 1997). The contribution of ILDs, however, has been incorporated in a number of recent models (Gaik, 1993; Breebaart *et al.*, 2001; Braasch and Blauert, 2003; Faller and Merimaa, 2004). It is thus hard to predict what would happen if binaural cues for the PE compete with each other, as a function of signal duration, and for temporally modulated broadband sounds like speech.

B. Onset and duration effects

The PE was generally observed for brief vocoded noise stimuli while it was more likely to fail with the longer speech sounds. Vocoding replaced the low-pass and wide-band noise stimuli with similar stimuli containing different samples of noise. Since processing was done in frequency bands which were roughly as wide as auditory filters, the spectral contour of the excitation pattern is minimally affected. However, the processed noise contains a different fine structure which is unrelated to that of the original lead–lag stimuli and does not carry the specific phase relationships that stem from adding lead and lag. The relative contribution of the fine structure to the PE depends on stimulus duration: For short noise bursts or clicks ($\leq$10 ms), onset information dominates while TFS information becomes more important for longer duration stimuli (Tobias and Schubert, 1959; Freyman and Zurek, 1997; Stecker and Hafter, 2002; Dizon and Colburn, 2006). The carrier has less influence for short bursts because the output of the auditory filter is dominated by its impulse response and the lead dominates the auditory nerve response because of peripheral compression and neural effects such as refractoriness (Hartung and Trahiotis, 2001).

The relative salience of the onset also increases when its slope is steeper (e.g., Rakerd and Hartmann, 1986). If one was to assume that the envelope of the vocoded speech sounds carries insufficient information for fusion and localization dominance after the onset, the lack of a pronounced onset with some words might be related to the failure of fusion. The onset of the word "teak" was about a factor of ten shorter than the onset of the other words—"teak" was the only word for which fusion and localization dominance appeared at short delays after vocoding. The effect of onset slopes on ETs, i.e. the maximum delay for fusion, has been described by Miller *et al*. (2009) for words.

Studies of lateralization discrimination with CI patients using low-rate electrical pulse trains showed good ITD sensitivity for the lead and degraded sensitivity in the lag at brief delays of 1–2 ms (van Hoesel, 2007; Agrawal, 2008). This indicates that the PE can operate for single lead–lag pulses under direct electrical stimulation of the auditory nerve and is in agreement with the results here for noise bursts. Using the paradigm of experiment 1, Seeber and Hafter (2007) demonstrated localization dominance for the wide-band noise burst and for the word "shape" in some listeners with clinical CIs. However, unlike in the present study, for the speech token, the majority of CI listeners indicated an image centered between the lead and lag location, even at long delays.

There may be an alternative explanation. The speech stimuli are longer than the noise bursts, thus allowing the listener to accumulate more information about the presence of less salient auditory objects like the lag. In subsequent informal listening tests, we compared the PE for trains of noise bursts with overall duration more similar to that of the speech sounds (500 ms duration). The lag was not audible at 1 ms delay in a pulse train with pulses of 5 ms duration at 10 Hz rate. For 10 ms pulse duration the lag was just audible in the pulse train, but essentially inaudible when a single pulse was presented. For 30 ms pulse duration the lag was audible for single and repeated presentations. Further study would be needed to confirm these observations, but they suggest that information about the lag could be integrated in a "multiple looks" strategy as well as from TFS with long pulse durations (Viemeister and Wakefield, 1991; Freyman and Zurek, 1997). Both could potentially explain why the PE tended to be stronger for vocoded noise bursts than for the longer speech stimuli.

C. Influence of TFS and interaural coherence

The interaural correlation of the carriers was varied in order to study the role of binaural information in the fine structure. Consistent ITDs do not exist in the purely random fine structure of an interaurally uncorrelated noise and the degree of correlation should affect the salience of fine-structure ITDs. As previously noted, the perceived width of the auditory image also changes with correlation (Blauert and Lindemann, 1986). The location of auditory objects is thought to be formed by an across-frequency integration of channel-wise position information derived from binaural cues (e.g., Stern *et al.*, 1988; Blauert, 1997). In the present study the correlated TFS carried consistent ITDs of zero, relevant for channels below about 1200 Hz. If this binaural information had been perceived independently of the envelope, a separate image should have appeared in the center of the head. Cluster analysis was used to determine if lateralization results were distributed as one or two images. The results for the words give evidence for both: Some subjects seemed to respond with a single image at short delays while 1/3 to 2/3 of the subjects, depending on condition, responded with split images even at delays as short as 2 ms. Lateralization results for the lead and lag images also show evidence of split images, one on the lead and one on the lag side, indicating that binaural information in the envelope affected perception. Envelope information appeared more salient for lower noise correlations, as the images were lateralized further from the head center. While previous PE studies showed the dominance of TFS for longer duration stimuli (Freyman and Zurek, 1997), the absence of a centered image for some subjects with the vocoding suggests that this is not the case. As listeners lateralized the images toward the sides, they appear to be created by binaural information in the envelope or, at least, by a combination of TFS and envelope information. The correlated TFS with its ITD of zero seems to pull the images toward the center.

The idea that correlated TFS is the glue that keeps information in the two ears bound together fails frequently in the present experiment. While the reason is not entirely clear, one contributing factor might be the inconsistency of information in TFS and envelope. With narrow peripheral filtering, envelope information can be "recovered" from the fine structure (Ghitza, 2001). The TFS of the carrier noise might have produced a "recovered" envelope that interacted with the envelope extracted from the lead–lag signal. If so, the envelope from the lead–lag signal might not be accurately transmitted to the auditory nerve, leading to a misrepresentation of the relation between lead and lag. For interaurally uncorrelated carriers, this misrepresentation could have reduced interaural envelope similarity in a way that interfered with the ability of the auditory system to group information across ears. This failure to bind signals between ears into a single percept would be analogous to what happens when listening to uncorrelated noise, where two images are localized at the ears. The decorrelation of the envelope might not be important for transient stimuli where the contrast between envelope and TFS information may not appear. In this regard, several studies have shown that brief lead and lag stimuli need not be correlated to evoke the PE (e.g., Shinn-Cunningham *et al.*, 1995; Yang and Grantham, 1997).

Another reason for the failure of precedence might stem from a lack of predictability of information across time. The fine structure at one time instant is uncorrelated with itself at a later point in time, because random noise carriers were used. The auditory system might not be able to attribute TFS information to the lead because it might expect a delayed copy of previously presented fine structure. This may explain why the PE fails even for correlated carriers.

D. A failure of auditory grouping?

Let us now explore an alternative explanation for the apparent failure of the PE. In discrimination suppression experiments the PE is seen as a binaural phenomenon where ITD thresholds in the lag are raised due to the presence of the lead (e.g., Litovsky *et al.*, 1999). In the present localization dominance experiment, instead, participants are instructed to indicate the position of one or two auditory objects. This requires auditory objects being formed and positioned in the auditory scene on the basis of available temporal, spectral, and binaural information. Viewing localization dominance experiments in the light of auditory scene analysis suggests that the fusion of a leading and a lagging sound into one perceptual event at short delays and their segregation into two events at longer delays might not be based solely on binaural information, but also on monaural grouping cues. Roberts *et al.* (2004) showed that ETs for brief sounds were surprisingly similar in monaural and binaural presentation, suggesting that the suppression of binaural information in the lag may have only a small influence on the perceptual fusion of lead and lag as evidenced in ETs. Further evidence comes from a neurophysiological study which showed similar time constants for recovery from lag suppression in both the horizontal and the median plane (Litovsky and Yin, 1998), suggesting that it is determined mostly by other than binaural cues. We suggest that in PE experiments where the perceptual fusion of lead and lag affects (localization) responses, the PE should be seen as part of auditory scene analysis in which monaural information such as onset slopes, spectral similarity, and pitch information is evaluated along with binaural stimulus parameters to determine fusion or segregation of lead and lag.

From the perspective of auditory scene analysis, another explanation for the failure of PE fusion might be possible: Lead and lag cannot be grouped into a single event because the lag cannot be identified as a copy of the lead. What information guides the auditory system to recognize the lag as a copy of the lead and not as some other sound? Several studies on auditory scene analysis have shown that spectral information and defined temporal onsets provide the strongest grouping cues (Cooke and Ellis, 2001).

By smearing the spectral content with broad analysis filters, vocoder processing destroys most of the fine spectral information useful, e.g., for perceiving the pitch of the test words. Hence fusion of lead and lag is not possible on the basis of spectral detail and the lack of this information might not be compensated for by other grouping information, such as common amplitude modulation.

Most studies of the PE have used stimuli with sharp transients (Litovsky *et al.*, 1999). However, transients provide a cue toward segregation if they are presented non-simultaneously (Bregman, 1990; Darwin and Ciocca, 1992). Due to its delayed onset, the lag should be easier to segregate from the lead if it carries a sharper onset, i.e. when it is temporally more distinct from the lead. Consistent with an interpretation in the context of auditory scene analysis, this seems to be the case: ETs are in the region of 3–8 ms for highly transient stimuli like pulse trains, while they are significantly larger for speech (20 ms) (Blauert, 1997; Stecker and Hafter, 2002).

We conclude that in order to explain all aspects of the PE in localization dominance experiments it must be seen in the larger context of auditory scene analysis, in which segregation on the basis of monaural information plays a role alongside the contribution from binaural cues.

VIII. CONCLUSIONS

The PE, that is the localization dominance of a leading sound over a lagging sound, was studied with lead–lag stimuli passed through a binaural noise-band vocoder. For brief stimuli the PE was generally active despite the replacement of TFS of the lead–lag sounds with that of noise carriers. For longer duration speech stimuli the PE was more likely to fail such that two images were lateralized, one toward the lead and one toward the lag, even for delays as short as 2 ms. Both images appeared more centered the higher the interaural correlation of the noise carriers, but in several subjects the split into two images remained even for fully correlated carriers. The classical duplex theory cannot explain this split as the ITD in the TFS would suggest a single image in the head center. As split images were somewhat lateralized, they seem to follow binaural information in ILDs and envelope ITDs separately for lead and lag. We suggest an alternative interpretation of the results in the context of auditory scene analysis: By replacing the fine structure with that of noise the auditory system loses the ability to fuse lead and lag into one object. This raises many new questions: What information is necessary to identify the lag as a copy of the lead? How is lead–lag fusion affected by grouping information? What role does the consistency of information between carrier and envelope play as well as that of interaural information? Does information need to be predictable over time to help identify the lag as a copy of the lead?

If we assume that results with a vocoder can guide our understanding of listening with CIs then the results suggest that the PE would fail with CIs because the carrier fine structure is replaced with information unrelated to the lead–lag stimuli. However, the prolonged exposure to fine structure unrelated to the sound, as in current CIs, might alter the relative weighting of TFS and envelope information, such that the former will be disregarded. In this case the PE might be possible with CIs, as it could be evoked here for brief, vocoded stimuli on the basis of envelope information.

Based on the new hypothesis that grouping information affects fusion of lead and lag in localization dominance experiments, more work will be needed to determine whether a lack of accurate fine-structure representation can be alleviated by enhancing other cues for grouping between lead and lag to re-gain the PE with vocoders and with CIs.

ACKNOWLEDGMENTS

We thank the subjects for their participation and Ruth Litovsky, Wes Grantham, Ian Wiggins, and one anonymous reviewer for their suggestions toward improving this manuscript. Oliver Zobay helped with the statistical analysis (ANOVAs). This work was funded by National Institutes of Health RO1 DCD 00087 and the first author was supported by the Intramural Programme of the Medical Research Council (UK) during data analysis and write-up of the manuscript.

Agrawal, S. (**2008**). "Spatial hearing abilities in adults with bilateral cochlear implants," Doctoral thesis, University of Wisconsin, Madison.

Blauert, J. (**1997**). *Spatial Hearing* (MIT Press, Cambridge, MA), 494 p.

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

Blauert, J., and Lindemann, W. (**1986**). "Spatial mapping of intracranial auditory events for various degrees of interaural coherence," J. Acoust. Soc. Am. **79**, 806–813.

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.

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.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA), 773 p.

Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Cooke, M., and Ellis, D. P. W. (**2001**). "The auditory organization of speech and other sources in listeners and computational models," Speech Commun. **35**, 141–177.

Cullington, H., and Zeng, F. G. (**2008**). "Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects," J. Acoust. Soc. Am. **123**, 450–461.

Darwin, C. J., and Ciocca, V. (**1992**). "Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component," J. Acoust. Soc. Am. **91**, 3381–3390.

Dizon, R. M., and Colburn, H. S. (**2006**). "The influence of spectral, temporal, and interaural stimulus variations on the precedence effect," J. Acoust. Soc. Am. **119**, 2947–2964.

Dudley, H. (**1939**). "Remaking speech," J. Acoust. Soc. Am. **11**, 169–177.

Edmonds, B. A., and Culling, J. F. (**2005**). "The role of head-related time and level cues in the unmasking of speech in noise and competing speech," Acta Acust. Acust. **91**, 546–553.

Faller, C., and Merimaa, J. (**2004**). "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," J. Acoust. Soc. Am. **116**, 3075–3089.

Freyman, R. L., and Zurek, P. M. (**1997**). "Onset dominance in lateralization," J. Acoust. Soc. Am. **101**, 1649–1659.

Fu, Q. J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Gaik, W. (**1993**). "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," J. Acoust. Soc. Am. **94**, 98–110.

Garadat, S. N., Litovsky, R. Y., Yu, G., and Zeng, F. G. (**2009**). "Role of binaural hearing in speech intelligibility and spatial release from masking using vocoded speech," J. Acoust. Soc. Am. **126**, 2522–2535.

Ghitza, O. (**2001**). "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," J. Acoust. Soc. Am. **110**, 1628–1640.

Hafter, E. R., and Buell, T. N. (**1990**). "Restarting the adapted binaural system," J. Acoust. Soc. Am. **88**, 806–812.

Hafter, E. R., and Jeffress, L. A. (**1968**). "Two-image lateralization of tones and clicks," J. Acoust. Soc. Am. **44**, 563–569.

Hartmann, W. M. (**1983**). "Localization of sound in rooms," J. Acoust. Soc. Am. **74**, 1380–1391.

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.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.

Hopkins, K., and Moore, B. C. J. (**2009**). "The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise," J. Acoust. Soc. Am. **125**, 442–446.

Houtgast, T., and Aoki, S. (**1994**). "Stimulus-onset dominance in the perception of binaural information," Hear. Res. **72**, 29–36.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Gunzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Litovsky, R. Y., Parkinson, A., and Arcaroli, J. (**2009**). "Spatial hearing and speech intelligibility in bilateral cochlear implant users," Ear Hear. **30**, 419–431.

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., and Yin, T. C. (**1998**). "Physiological studies of the precedence effect in the inferior colliculus of the cat. II. Neural mechanisms," J. Neurophysiol. **80**, 1302–1316.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound revisited," J. Acoust. Soc. Am. **111**, 2219–2236.

Miller, S. D., Litovsky, R. Y., and Kluender, K. R. (**2009**). "Predicting echo thresholds from speech onset characteristics," J. Acoust. Soc. Am. Express Lett. **125**, EL134–EL140.

Nábělek, A. K., and Robinson, P. K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**, 1242–1248.

Nuetzel, J. M., and Hafter, E. R. (**1981**). "Discrimination of interaural delays in complex waveforms: Spectral effects," J. Acoust. Soc. Am. **69**, 1112–1118.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Rakerd, B., and Hartmann, W. M. (**1986**). "Localization of sound in rooms, III: Onset and duration effects," J. Acoust. Soc. Am. **80**, 1695–1706.

Roberts, R. A., Koehnke, J., and Besing, J. (**2004**). "Effects of reverberation on fusion of lead and lag noise burst stimuli," Hear. Res. **187**, 73–84.

Schön, F., Müller, J., and Helms, J. (**2002**). "Speech reception thresholds obtained in a symmetrical four-loudspeaker arrangement from bilateral users of MED-EL cochlear implants," Otol. Neurotol. **23**, 710–714.

Seeber, B. (**2002**). "A new method for localization studies," Acta Acust. Acust. **88**, 446–450.

Seeber, B., Baumann, U., and Fastl, H. (**2004**). "Localization ability with bimodal hearing aids and bilateral cochlear implants," J. Acoust. Soc. Am. **116**, 1698–1709.

Seeber, B. U., and Fastl, H. (**2003**). "Subjective selection of non-individual head-related transfer functions," in *Proceedings of the 9th International Conference on Auditory Display*, edited by E. Brazil and B. Shinn-Cunningham (Boston University Publications Production Department, Boston), pp. 259–262.

Seeber, B., and Fastl, H. (**2008**). "Localization cues with bilateral cochlear implants," J. Acoust. Soc. Am. **123**, 1030–1042.

Seeber, B., and Hafter, E. (**2007**). "Breakdown of precedence with cochlear implants—Simulations show importance of spectral cues," in *Proceedings of the 2007 Conference on Implantable Auditory Prostheses*, July 15–20, Granlibakken Conference Grounds, Lake Tahoe, CA, p. 210.

Seeber, B. U., Kerber, S., and Hafter, E. R. (**2010**). "A system to simulate and reproduce audio-visual environments for spatial hearing research," Hear. Res. **260**, 1–10.

Shannon, R. V., Fu, Q. J., and Galvin, J. (**2004**). "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation," Acta Oto-Laryngol., Suppl. **552**, 50–54.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Shinn-Cunningham, B. G., Kopco, N., and Martin, T. J. (**2005**). "Localizing nearby sound sources in a classroom: Binaural room impulse responses," J. Acoust. Soc. Am. **117**, 3100–3115.

Shinn-Cunningham, B., Zurek, P. M., Durlach, N. I., and Clifton, R. K. (**1995**). "Cross-frequency interactions in the precedence effect," J. Acoust. Soc. Am. **98**, 164–171.

Stecker, G. C., and Hafter, E. R. (**2002**). "Temporal weighting in sound localization," J. Acoust. Soc. Am. **112**, 1046–1057.

Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (**1988**). "Lateralization of complex binaural stimuli: A weighted-image model," J. Acoust. Soc. Am. **84**, 156–165.

Tobias, J. V., and Schubert, E. D. (**1959**). "Effective onset duration of auditory stimuli," J. Acoust. Soc. Am. **31**, 1595–1605.

Tollin, D. J., and Henning, G. B. (**1999**). "Some aspects of the lateralization of echoed sound in man. II. The role of the stimulus spectrum," J. Acoust. Soc. Am. **105**, 838–849.

van Hoesel, R. J. M. (**2007**). "Sensitivity to binaural timing in bilateral cochlear implant users," J. Acoust. Soc. Am. **121**, 2192–2206.

Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). "The precedence effect in sound localization," Am. J. Psychol. **LXII**, 315–337.

Wilson, B. S., and Dorman, M. F. (**2008**). "Cochlear implants: A remarkable past and a brilliant future," Hear. Res. **242**, 3–21.

Yang, X., and Grantham, D. W. (**1997**). "Cross-spectral and temporal factors in the precedence effect: Discrimination suppression of the lag sound in free-field," J. Acoust. Soc. Am. **102**, 2973–2983.

Zeng, F. G., Nie, K., Liu, S., Stickney, G., Rio, E. D., Kong, Y. Y., and Chen, H. (**2004**). "On the dichotomy in auditory perception between temporal and fine structure cues (L)," J. Acoust. Soc. Am. **116**, 1351–1354.

Zurek, P. M. (**1980**). "The precedence effect and its possible role in the avoidance of interaural ambiguities," J. Acoust. Soc. Am. **67**, 952–964.

Zurek, P. M. (**1987**). "The precedence effect," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer, New York), pp. 85–105.

Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics, Facts and Models* (Springer, Berlin, Heidelberg, New York), 416 p.