Sound segregation based on temporal envelope structure and binaural cues

Othmar Schimmel
Eindhoven University of Technology, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

Steven van de Par[a] and Jeroen Breebaart
Philips Research, High Tech Campus 36, NL-5656 AE Eindhoven, The Netherlands

Armin Kohlrausch
Eindhoven University of Technology, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands
and Philips Research, High Tech Campus 36, NL-5656 AE Eindhoven, The Netherlands

(Received 29 January 2007; revised 18 April 2008; accepted 21 May 2008)

The ability to segregate two spectrally and temporally overlapping signals based on differences in temporal envelope structure and binaural cues was investigated. Signals were a harmonic tone complex (HTC) with 20 Hz fundamental frequency and a bandpass noise (BPN). Both signals had interaural differences of the same absolute value, but with opposite signs to establish lateralization to different sides of the medial plane, such that their combination yielded two different spatial configurations. As an indication for segregation ability, threshold interaural time and level differences were measured for discrimination between these spatial configurations. Discrimination based on interaural level differences was good, although absolute thresholds depended on signal bandwidth and center frequency. Discrimination based on interaural time differences required the signals' temporal envelope structures to be sufficiently different. Long-term interaural cross-correlation patterns or long-term averaged patterns after equalization-cancellation of the combined signals did not provide information for the discrimination. The binaural system must, therefore, have been capable of processing changes in interaural time differences within the period of the harmonic tone complex, suggesting that monaural information from the temporal envelopes influences the use of binaural information in the perceptual organization of signal components.

© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2945159]

PACS number(s): 43.66.Pn, 43.66.Mk, 43.66.Rq [RYL] Pages: 1130–1145

I. INTRODUCTION

The perceptual organization of sound sources within an auditory scene is based on grouping the signal components that are likely to come from the same source into one auditory object. The grouping of signal components into auditory objects is established by similarities in spectral, temporal, and/or spatial cues, whereas dissimilarities in these cues cause auditory objects to segregate from each other (cf. Bregman, 1990). The process by which a composite signal is analyzed to identify its constituent components, and the interactions between various grouping cues that contribute to segregation, are, however, not yet fully understood.

For example, when the spectra of two simultaneously presented auditory objects are different, the binaural information within each left-right pair of auditory filters is dominated by the spatial cues of the auditory object that has the most energy in the frequency range of those auditory filters. Although, in principle, this could provide sufficient cues for parsing individual auditory objects, two concurrent tones that are harmonically related cannot be segregated by their differences in interaural timing (Buell and Hafter, 1991). This suggests that signal components with a harmonic relationship are grouped into one auditory object, regardless of their spatial cues. Furthermore, grouping of competing frequency bands to form vowel-like sounds cannot be performed based solely on shared interaural time differences (Culling and Summerfield, 1995). Thus, in addition to different spectral cues and interaural time differences, additional evidence supporting the presence of another auditory object is required for segregation to occur. Such evidence could consist of a repeating tone, which can capture one harmonic of a vowel into a separate auditory object (Hukin and Darwin, 1995; Darwin and Hukin, 1997). Alternatively, correlated dynamic variation in frequency and/or amplitude, such as in speech, is sufficient to segregate two vowel-like sounds based on interaural time differences (Stern *et al.*, 2006) or increases speech intelligibility in the presence of spatially separated competing speech or noise signals (Noble and Perrett, 2002).

The above studies indicate that segregation of concurrent auditory objects depends on the specific spectral, temporal, and spatial cues of the signal components. In contrast to spectral and temporal grouping cues, spatial cues alone do not appear to be sufficient to *cause* segregation (cf. Culling and Summerfield, 1995). Spatial cues do, however, *add to* segregation that is established on the basis of monaural spectral or temporal grouping cues (Shackleton and Meddis, 1992; Darwin and Hukin, 1998). Subjects were shown to use the continuity of interaural time differences as an important

[a]Electronic mail: steven.van.de.par@philips.com

cue for sequential organization of signal components, i.e., tracking an auditory object across time (Hukin and Darwin, 1995; Darwin and Hukin, 1997, 1999; Shinn-Cunninham, 2005). It seems that monaural grouping cues are essential for segregation of spatially separated auditory objects to occur, and that these monaural cues influence the contribution of binaural cues to the segregation.

Based on these results from literature, the question arises how the monaural grouping cues influence the ability to use binaural grouping cues in the perceptual organization of concurrent signals. The current study focused on the relationship between temporal and binaural cues. Experiments that explored whether and to what extent subjects can discriminate between the different spatial configurations of two concurrent and spectrally overlapping signals that differed in their temporal envelope structures are described. Each of the two signals is easily recognized when listened to in isolation. By presenting the signals simultaneously, with equal spectral excitation patterns and interaural differences of the same absolute value but with opposite signs, binaural cues from both signals are, on average, equally represented in each frequency band of the signals. This way, only temporal cues from the signals' different temporal envelope structures and binaural cues are available in the composite signal for discrimination between its different spatial configurations. It is of interest to investigate whether subjects can analyze the temporal envelope cues and the interaural differences present in the composite signal to distinguish between the different spatial configurations, and to what extent they are able to associate the binaural information with the underlying signal components.

II. GENERAL METHOD

As an indication for segregation ability, the experiments explored the ability of subjects to discriminate between the spatial configurations of two simultaneously presented, spectrally overlapping signals with different temporal envelopes. The two signals were a harmonic tone complex (HTC) with a 20 Hz fundamental frequency and a bandpass noise (BPN), which occupied the same spectral range. Both signals had interaural differences of the same absolute value, but with opposite signs, to establish lateralization to different sides of the medial plane. Their combination yielded two different spatial configurations, i.e., a configuration with the HTC on the one side and the BPN on the other side, and the reverse configuration.

All experiments used the same method to measure thresholds for various signal parameters. A three-interval, three-alternative, forced-choice procedure with an adaptive parameter adjustment was used in a within-subject experimental design with counterbalanced block randomization. The subjects' task was to identify the interval that was different from the other two intervals. Feedback was provided after each trial. The adaptive parameter [interaural time difference (ITD) or interaural level difference (ILD)] was adjusted according to a two-down one-up rule, to track the 70.7%-correct response level (Levitt, 1971), by multiplying or dividing it with a certain factor. Initially, this factor was 2.51 ($=10^{8/20}$). After each second reversal, the factor was reduced by taking its square root until the minimum factor of 1.12 ($=10^{1/20}$) was reached. Another eight reversals were measured at this minimum factor, and the median of these eight values was used to estimate the threshold.

For each condition, at least four attempts were made by each subject to measure a threshold. However, when the adaptive interaural difference exceeded a limit of 2 ms ITD or 96 dB ILD, the tracking procedure was terminated and no threshold was registered. For conditions where incidently no threshold was registered, the measurements were repeated to obtain a total of at least four threshold values. Because of possible lateralization ambiguities in the ITD conditions, resulting from phase shifts beyond π for the highest-frequency components of the HTC due to the 1 ms ITD starting value, the conditions that did not yield threshold values were repeated with a 100 μs ITD starting value. For conditions that did not consistently yield threshold values, measurements were stopped and occasionally obtained threshold values were discarded. The measured thresholds for each condition were pooled and checked for severe outliers (thresholds that deviated from the average more than three times the interquartile range of the pooled data for each condition), which were then removed from the data set. Five male subjects without any reported hearing problem, including the four authors, participated in the experiments.

The experiments were conducted in an acoustically isolated listening room at the Philips Research Laboratories. A computer running MATLAB software generated the stimuli and automated the experiments and data collection. Digital stimuli were converted to analog signals by a Marantz CDA-94 two-channel 16 bit digital-to-analog converter at a sampling rate of 44.1 kHz and presented to the subjects over Beyer Dynamic DT990 Pro headphones.

In Sec. III, the ability to discriminate between the spatial configurations of two spectrally and temporally overlapping signals with different temporal envelope structures is explored. In Secs. IV and V, it is checked whether the results of Sec. III may have been obtained by judging an overall lateralization of the composite signal or by monaural listening. In Sec. VI, various bandwidth conditions were explored to further investigate the observed influence of bandwidth on discrimination performance in Sec. III. In Sec. VII, various temporal envelope structures were applied to the individual signals to investigate their effect on the ability to discriminate between the spatial configurations of the signals.

III. EXPERIMENT 1

The first experiment investigated the extent to which subjects could discriminate between the spatial configurations of two spectrally and temporally overlapping signals with different temporal envelope structures. As a reference for discrimination between the spatial configurations of the signals, the just-noticeable differences in lateralization for each of the individual signals were also measured.

A. Stimuli

The stimuli consisted of two signals, for which the choice of parameters was inspired by a related earlier study (van de Par *et al.*, 2005): a HTC with a 20 Hz component spacing and a BPN. The HTC was defined as a complex of sinusoidal components at multiples of 20 Hz, with zero starting phase as follows

$$x(n) = \sum_{k=i}^{j} A \sin\left(2\pi k f_0 \frac{n}{f_s}\right), \qquad (1)$$

where n is the sample number, f_s is the sample frequency, and i and j are integers that indicate the harmonic number of the lowest and highest components in the HTC. The noise was generated in the time domain by creating a 3000 ms buffer containing a broadband noise with a Gaussian probability distribution. This buffer was transformed to the frequency domain using a fast fourier transform (FFT), and the appropriate samples were set to zero to obtain the required bandwidth and center frequency. Using an inverse FFT, the buffer was transformed back to the time domain. For each threshold measurement, one bandpass filtered buffer was generated, and for each trial, three independent 400 ms excerpts were selected from the buffer by drawing random starting positions from a uniform distribution. Both signals had the same spectral range with a flat spectral envelope and the same overall level [65 dB sound pressure level (SPL)], but differed in their temporal envelope structures. Whereas the BPN had an irregular temporal envelope, the envelope of the HTC was highly modulated and had a period of 50 ms.

For measuring the discrimination between their spatial configurations, the two signals were presented concurrently, resulting in an overall stimulus level of 68 dB SPL. The target interval of the forced-choice procedure had interaural time or level differences for the two signals such that the HTC was lateralized to the right and the BPN to the left, using interaural time or level differences of the same absolute value, but with opposite signs. In the two reference intervals, the interaural differences of both signals were reversed. The subjects' task was to identify the target interval that differed from the other two intervals, i.e., the interval in which the HTC was lateralized to the right and the BPN to the left. These conditions are referred to as *composite signal* conditions.

For measuring the just-noticeable differences in lateralization for the individual signals, the same HTC and BPN were presented in isolation. The target interval of the forced-choice procedure had an interaural time or level difference such that the signal was lateralized to the right. The reference intervals had an identical opposite binaural cue such that the signal was lateralized to the left. Again, the subjects' task was to identify the target interval that differed from the other two intervals, i.e., the interval in which the signal was lateralized to the right. These conditions are referred to as *single signal* conditions. Please note that the interaural differences as reported hereafter are relative to the medial plane, and that, due to the lateralizations to opposite sides of the medial plane, the total interaural difference cue between the signals is, in fact, twice the size of these interaural differences.

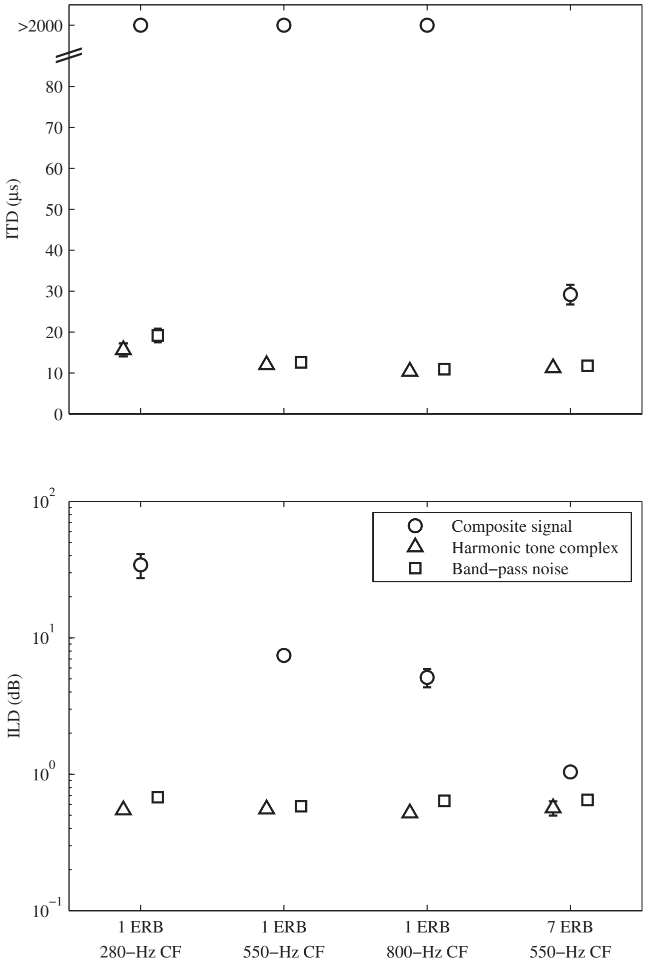

FIG. 1. Mean ITD (top panel) and ILD (bottom panel) thresholds for the composite and single signal conditions from Exp. 1. ITD thresholds that could not be measured are indicated by a symbol at a threshold of >2000 μs. Error bars represent the standard errors of the mean. Error bars smaller than the symbol size are omitted. Please remember that the total interaural difference cue between the two signals is, in fact, twice the size of these interaural differences.

The two signals were of various equal bandwidths. In the narrowband conditions, the signals were the width of one auditory filter [1 Equivalent Rectangular Bandwith (ERB)] and centered at 280 Hz (bandwidth of 60 Hz), 550 Hz (bandwidth of 80 Hz), or 800 Hz (bandwidth of 100 Hz). In the wideband condition, the signals were 600 Hz wide (7 ERB) and centered at 550 Hz. The intervals had a duration of 400 ms, including 30 ms raised-cosine onset and offset ramps to avoid spectral splatter, and were separated by 300 ms of silence. For the ILD conditions, the level changes were such that the level of the combined signals remained constant at 68 dB SPL.

B. Results

Figure 1 displays the mean thresholds and the standard errors of the mean of the pooled data of the five subjects for the composite and single signal conditions. The top panel shows the data for the ITD conditions and the bottom panel the data for the ILD conditions. The abscissa indicates the four signal bandwidth conditions. The ordinate indicates the size of the thresholds in microseconds (ITD) or decibels

(ILD). The symbols represent the mean thresholds for the composite signal conditions (circles), and for the HTC (triangles) and BPN (squares) in the single signal conditions. The ITD conditions for which no threshold could be obtained, because the adjusted value exceeded the procedural limit of 2 ms, are indicated by a symbol at a threshold of >2000.

For the ITD conditions, all thresholds for the single signal conditions were about 10–20 μs, and decreased slightly for both signals with an increase in center frequency. The thresholds for the wideband condition agreed with those for the narrowband condition with the highest center frequency. These results indicate similar lateralization performance for both signals when presented in isolation, independent of the signals' detailed spectral and temporal features. For the composite signal conditions, no thresholds could be obtained for the narrowband conditions, indicating that discrimination between the two signals' spatial configurations based on interaural time differences was not possible when their spectral energy was limited to one auditory filter. For the wideband condition, the threshold was 29 μs, indicating that such a discrimination was possible when the spectral energy of the two signals was present across multiple auditory filters.

For the ILD conditions, all thresholds for the single signal conditions were about 0.6 dB, independent of the center frequency and bandwidth. For the composite signal conditions, the thresholds were in the range 1–32 dB. For the narrowband conditions, thresholds decreased from 32 to 5 dB with increasing center frequency. The 32 dB threshold for the narrowband condition at 280 Hz center frequency is, however, beyond a plausible range for naturally occurring interaural level differences, and reflects the difficulty of assigning these binaural cues to either one of the two signals. For the wideband condition, the threshold of 1 dB was close to the corresponding thresholds in the single signal condition. These results indicate that, in contrast to the narrowband ITD conditions, discrimination between the two signals' spatial configurations based on interaural level differences was possible within a single auditory filter, although the thresholds were considerably higher than for the single signal conditions. Analogous to the wideband ITD conditions, performance in the composite signal conditions was best when the signals' spectrum covered multiple auditory filters.

C. Discussion

When asked, all subjects reported that in the composite signal the two constituent signals, and predominantly the HTC, could easily be recognized and, especially in the ILD conditions, lateralized when they had a sufficiently large interaural difference cue. At an interaural difference cue close to the threshold, their individual lateralizations would become much harder to discern, although the presence of the HTC, as the focus for identifying the target interval, could still be recognized in the composite signal. This recognition may, however, be due to the extended exposure to, and focus on the HTC during the procedure, and it is possible that both signals would otherwise have been merged into one auditory object based on their common properties. It seems that in these conditions, which are (close to) being perceived as diotic, the known characteristics of the specific temporal envelope of the HTC still allowed its segregation from the composite signal.

The fact that subjects were able to discriminate between the spatial configurations of the target and reference intervals in the composite signal condition suggests that subjects were able to segregate the two signals and discern the lateralization of at least one of them. A possible objection to this interpretation of the obtained results may be that successful identification of the target and reference intervals also could have been performed otherwise, for instance, based on an asymmetry in the overall spatial image of the composite signal.

The stimuli were physically constructed in such a way that, due to the two constituent signals' interaural differences of the same absolute value but with opposite signs, the long-term interaural cross-correlation patterns or long-term patterns after equalization-cancellation for the composite signal were essentially identical for target and reference intervals. Figure 2 shows an example of the long-term interaural cross-correlation patterns for the ITD condition of the individual wideband signals (dotted and dashed lines) and their composite signal (solid line) for one interval, computed from the output of the auditory filter centered at 550 Hz of a gammatone filterbank. In this example, the individual signals have an opposite interaural time difference of 100 μs, which is reflected in slightly different positions of the maxima of the cross-correlation patterns. The cross-correlation pattern of the composite signal has its maximum at 0 μs, and is highly symmetrical, as can be seen by comparing the pattern (the solid line) with its mirrored version (the circles on the solid line). Therefore, the long-term interaural cross correlation of the composite signal is not expected to provide directional information for distinguishing between target and reference intervals.

Similarly, the spatial perception of a single auditory object consisting of multiple merged signals is based on a weighted average of the various directional cues of the constituent signals, which fuse into a single intracranial image (Stellmack and Lutfi, 1996). Given their common temporal onset, spectral cues, and identical but opposite interaural differences, an unsuccessful segregation of the two signals would be expected to result in lateralization in the center and, therefore, inhibit discrimination between the different spatial configurations of the target and reference intervals.

However, this reasoning is based on linear processing in the auditory system, while the contributions of the two different signals to the internal long-term interaural cross correlation may be different. For instance, peripheral compression may have reduced the peaks in the temporal envelope of the HTC stronger than those in the temporal envelope of the BPN. If so, the combination of the two signals could have been perceived as a single auditory object with an averaged lateralization different from the center. Then, the intracranial image would be lateralized to different sides for the target and reference intervals, and subjects could have used this change in overall lateralization of the composite signal to

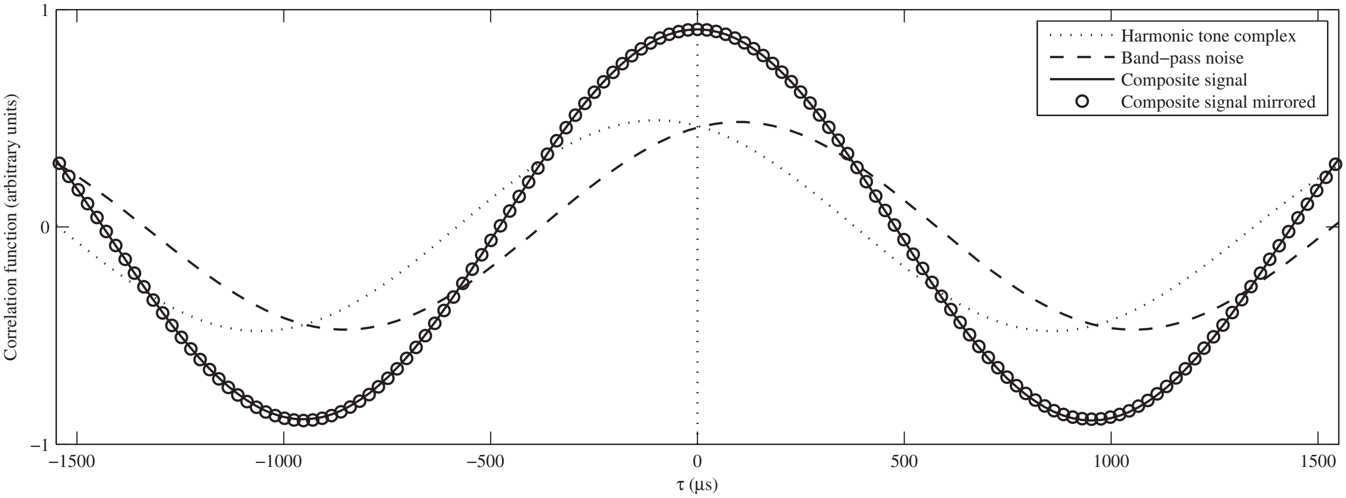

FIG. 2. Long-term interaural cross correlations of the signals at the output of the auditory filter centered at 550 Hz. The dash-dotted and dashed lines represent the interaural cross correlations of the individual signals, showing opposite lateralizations (peaks centered at $\tau = \pm 100$ μs). The solid line represents the interaural cross correlation of the composite signal, showing the effective cancellation of the individual signals' lateralizations (peak centered at $\tau = 0$ μs). The circles represent the cross correlation of the composite signal when plotted mirrored. Note that their pattern is nearly identical to the solid line.

identify the intervals, without the need to discriminate between the spatial configurations of the two signals.

Another possible criticism to the current interpretation of the results is that, in the ILD conditions, differences in the relative levels of the constituent stimulus components could have allowed monaural listening. The discrimination between the spatial configurations of the signals could then have been performed by listening to one ear and selecting the target interval based on overall monaural level cues. Because the interpretation of the results is only valid if the identification of the target and reference intervals could not have been performed other than through segregation and lateralization of at least one of the individual signals, Exps. 2A and 2B in the next section were designed to test the two mentioned alternative accounts for the results of Exp. 1.

IV. EXPERIMENT 2A

This first control experiment investigated whether subjects could have performed the identification of the target and reference intervals for the stimuli with interaural time differences of Exp. 1 by judging a possible overall lateralization of the composite signal, without the need to discriminate between the spatial configurations of the two signals. To effectively inhibit an overall lateralization as a cue for performing the identification, in each interval the actual interaural time difference, as adjusted by the adaptive procedure, was changed for both signals by the same amount using a random offset (rove).

A. Stimuli

The measurements were limited to the wideband ITD condition of Exp. 1, using the same HTC and BPN in the composite signal condition. The amount of roving was random for each interval and equally distributed between zero and the size of the currently adapted interaural time difference, in order to enable a maximum rove while keeping the signals on the opposite sides of the medial plane. It was applied in the *same direction* to both signals to keep the absolute difference in directional cues between them constant. This way, a possible overall lateralization would change randomly in each interval within a trial, while the interaural difference cue between the signals was maintained.

B. Results

The mean threshold for the nonroving condition from Exp. 1 was 29 μs, and for the roving condition from the current experiment it was 34 μs. Both values had the same standard error of the mean of 2 μs. Analysis by a t test showed that these results were statistically not significantly different ($p = 0.109$).

C. Discussion

The 5 μs difference in mean thresholds between the nonroving and roving conditions was well below the 10–20 μs thresholds from the single signal conditions, indicating that the difference between the nonroving and roving conditions is too small to be considered perceptually relevant. If an overall lateralization of the composite signal was the main cue for identifying the target and reference intervals, a significant increase in mean threshold was expected. Finding no significant or perceptually relevant effect of roving the interaural differences on the mean threshold leads to the conclusion that the results from Exp. 1 cannot be explained by an overall lateralization of the composite signal.

V. EXPERIMENT 2B

This second control experiment investigated whether identification of the target and reference intervals for the stimuli with interaural level differences from Exp. 1 could have been performed by monaural listening, or whether it required an analysis of binaural cues. The contributions of monaural and binaural listening were addressed separately.

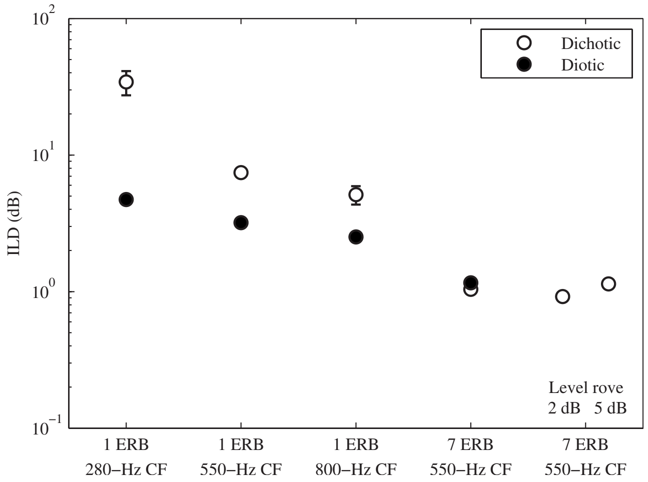

FIG. 3. Mean ILD thresholds for the dichotic signal conditions from Exp. 1 and the diotic and dichotic signal conditions from Exp. 2B. Error bars represent the standard errors of the mean. Error bars smaller than the symbol size are omitted. Please remember that the total interaural difference cue between the two signals is, in fact, twice the size of these interaural differences.

A. Stimuli

To investigate the potential of monaural listening, the measurements were repeated for all ILD conditions from Exp. 1, with the modification that only the right-ear signal was presented. It was presented diotically, to establish that all binaural cues were excluded and only the monaural level cues within one ear, resulting from the different temporal envelopes of the HTC and the BPN, were available for identifying the target and reference intervals.

To investigate the influence of binaural listening, the measurements were repeated for the wideband ILD condition from Exp. 1. For each interval within a trial, a random level rove, equally distributed between −2 and 2 dB or between −5 and 5 dB and sufficiently larger than the thresholds for the wideband composite and single signal conditions, was applied to the HTC. This level rove was followed by an independent level normalization of the composite left- and right-ear signals to 68 dB SPL each. Due to the resulting variability in the relative levels of the HTC and the BPN, the monaural level cues were unreliable and only the binaural cues were available for identifying the target and reference intervals.

B. Results

Figure 3 displays the mean thresholds and standard errors of the mean of the pooled data of the five subjects for the dichotic signal conditions from Exp. 1 and the diotic and dichotic signal conditions from the current experiment. The abscissa indicates the three narrowband conditions, the wideband condition, and the level-rove wideband condition. The ordinate indicates the size of the thresholds in decibels. The symbols represent the mean thresholds for the dichotic signal conditions (open markers) and the diotic signal conditions (closed markers).

Monaural (diotic) thresholds in the narrowband conditions decreased with an increase in center frequency, much like the corresponding binaural (dichotic) thresholds from Exp. 1. However, the monaural thresholds were considerably lower than the binaural thresholds, indicating better performance in the diotic conditions. The monaural and binaural thresholds for the wideband condition were similar, indicating equivalent performance. Analysis of variance on the means of the dichotic versus diotic conditions showed significant effects of signal bandwidth ($F_{(3,153)}=152.88$, $p<0.001$) and interaural condition ($F_{(1,153)}=82.72$, $p<0.001$), and a significant interaction between these two parameters ($F_{(3,153)}=23.82$, $p<0.001$). Tukey *post hoc* analysis revealed that, for all three narrowband conditions, the diotic conditions were significantly different from the dichotic conditions. For the wideband condition, the diotic and dichotic conditions were statistically identical.

The thresholds for the two dichotic conditions in which the level was randomly roved to inhibit overall level cues for identifying the target and reference intervals were also similar to the threshold of the wideband condition without a level rove, showing equivalent performance as well. Analysis of variance on the means of the wideband nonroving and the two roving conditions showed that these three conditions were statistically not significantly different from each other ($F_{(2,63)}=2.02$, $p=0.142$).

C. Discussion

The difference between diotic and dichotic listening in the narrowband conditions shows that better performance could have been achieved if subjects were able to listen to the signals at one ear only, which apparently they could not. Because the main difference between the diotic and dichotic listening conditions is the presence of binaural cues, this finding reveals that binaural cues actually *reduce* the performance for narrowband stimuli, while for wideband stimuli, performance is similar. This result suggests that, for narrowband stimuli, binaural processing cannot be ignored. Although this may seem surprising, it was shown before that subjects are sometimes unable to ignore information presented to one ear while performing a task that could be solved with information simultaneously presented to the other ear (Heller and Trahiotis, 1995; Gallun *et al.*, 2007). For the wideband conditions, the similarity between the results for diotic and dichotic listening indicates that the target and reference intervals could be identified equally well using either monaural or binaural cues, making it impossible to distinguish between monaural and binaural listening in the wideband ILD condition.

To investigate the contribution of binaural listening in the wideband ILD condition, the relative levels of the HTC and the BPN were randomly roved. Due to the random level roves, the overall monaural level cues were not reliable and identification of the target and reference intervals must have resulted from binaural processing. The results for these level-roving conditions were, however, similar to the corresponding nonroving condition of Exp. 1. This result suggests that, for the wideband ILD conditions, identification of the target

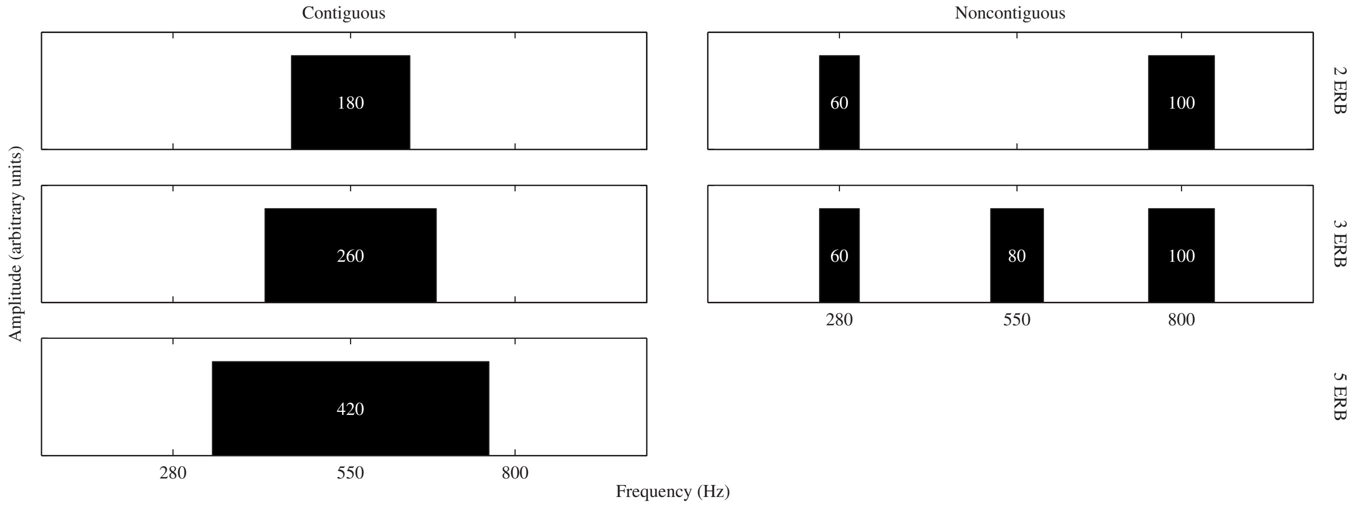

FIG. 4. Schematic illustration of the conditions of Exp. 3: Signal center frequency and bandwidth for the HTC and the BPN (in Hz), with the spectral energy distributed over contiguous or noncontiguous auditory filters.

and reference intervals could have been mediated by binaural cues. It remains, therefore, impossible to distinguish between monaural and binaural listening in these conditions.

VI. EXPERIMENT 3

The results of Exp. 1 showed that signal bandwidth has a pronounced effect on the ability to discriminate between the spatial configurations of the two signals. To investigate whether the observed increase in performance with an increase in bandwidth is due to across-channel or within-channel cues, various bandwidth conditions were explored, with the spectral energy of both signals distributed over either contiguous or noncontiguous auditory filters.

A. Stimuli

Both the HTC and the BPN were presented with bandwidths of 2, 3, and 5 ERB, all centered at 550 Hz. In the case of 2 and 3 ERB bandwidths, both signals were presented in contiguous as well as in noncontiguous auditory filters, i.e., the spectral energy of the signals was distributed over frequency bands of 1 ERB wide that were not adjacent. These noncontiguous conditions involved the same bands as the narrowband conditions from Exp. 1. For the 2 ERB condition, the bands were centered at 280 and 800 Hz; for the 3 ERB condition, the bands were centered at 280, 550, and 800 Hz. Figure 4 gives a schematic overview of the spectral conditions.

B. Results

Figure 5 displays the mean thresholds and the standard errors of the mean of the pooled data of the five subjects for the various signal bandwidth conditions. The top panel shows the data for the ITD conditions and the bottom panel shows the data for the ILD conditions. The abscissa indicates the five signal bandwidth conditions, including the narrowband centered at 550 Hz (1 ERB) and wideband (7 ERB) conditions from Exp. 1. The ordinate indicates the size of the thresholds in microseconds (ITD) or decibels (ILD). The symbols represent the mean thresholds for the contiguous auditory filter conditions (open markers) and the noncontiguous auditory filter conditions (closed markers).

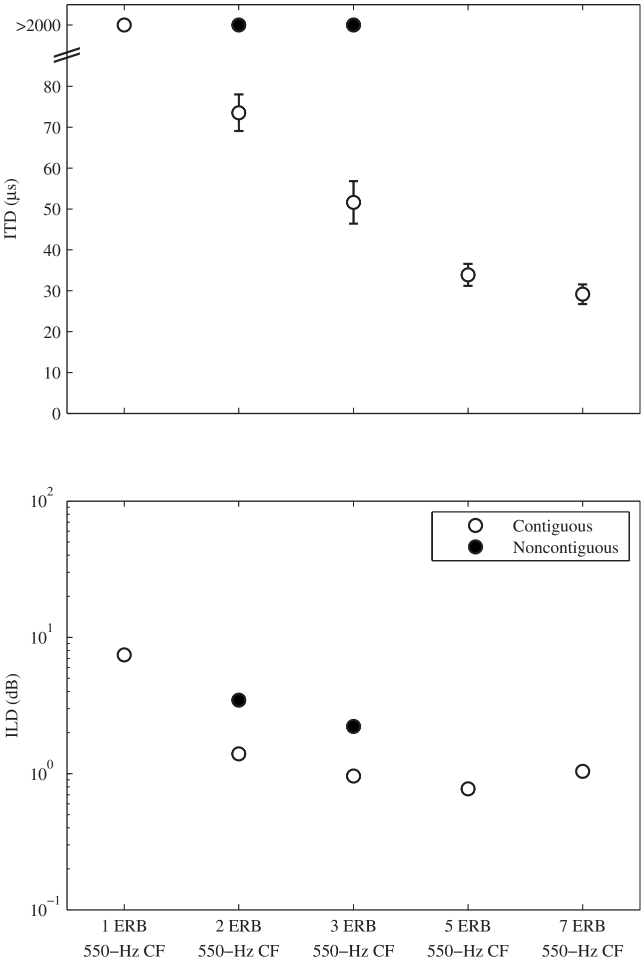

FIG. 5. Mean ITD (top panel) and ILD (bottom panel) thresholds for the signal bandwidth conditions from Exp. 3. Included for reference are the results for the narrowband (1 ERB) and wideband (7 ERB) conditions from Exp. 1. Error bars represent the standard errors of the mean. Error bars smaller than the symbol size are omitted. Please remember that the total interaural difference cue between the two signals is, in fact, twice the size of these interaural differences.

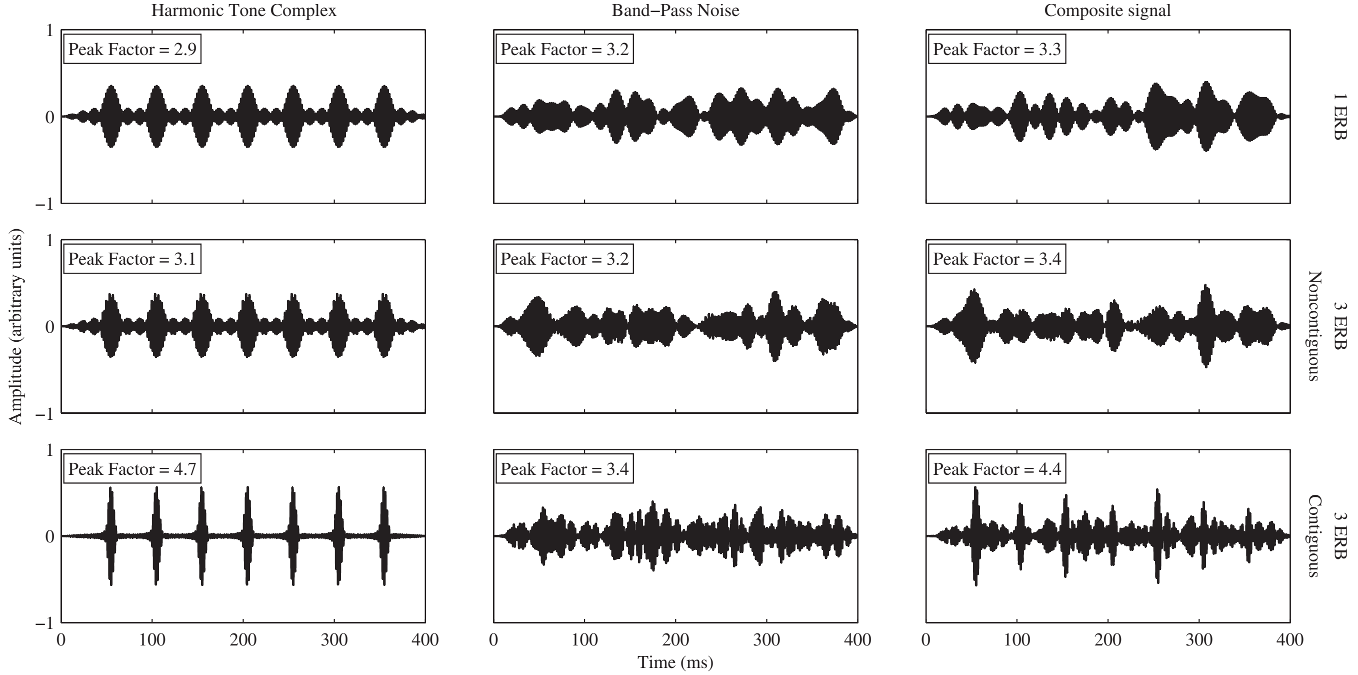

FIG. 6. Output of the auditory filter centered at 550 Hz for the HTC (left column), the BPN (middle column), and the composite signal (right column) at signal bandwidths of 1 ERB (top row) and 3 ERB in both noncontiguous (middle row) and contiguous auditory filters (bottom row). The insets show the peak or crest factor, i.e., the signal's peak amplitude divided by its rms value.

For the ITD conditions, thresholds of 74 and 52 μs were measured for the 2 and 3 ERB bandwidths, respectively, provided that the signals' spectral energy was in contiguous auditory filters. No thresholds could be obtained for these bandwidths when the signals' spectral energy was in noncontiguous auditory filters of 1 ERB width. These results indicate that the combination of the information across auditory filters *per se* did not improve the performance. The threshold of 34 μs for the 5 ERB condition was already similar to the previously established threshold (29 μs) for the 7 ERB wideband condition. These data indicate that discrimination between the signals' spatial configurations based on interaural time differences improves when their spectral energy covers an increasing number of contiguous auditory filters.

For the ILD conditions, the threshold for signals with a bandwidth of 2 ERB in contiguous auditory filters was 1.4 dB, and already very close to the thresholds of 1 dB for all larger signal bandwidths in contiguous auditory filters. For the signal bandwidths with the spectral energy in two and three noncontiguous auditory filters, thresholds were 3.5 and 2.2 dB, respectively 2.1 and 1.2 dB higher than for the corresponding bandwidth conditions with contiguous auditory filters. These data indicate that, again in contrast to the corresponding ITD conditions, discrimination of the signals' spatial configurations based on interaural level differences was possible for narrow bandwidths, even when the spectral energy was in noncontiguous auditory filters. When the bandwidth was increased by distributing the spectral energy over two or three noncontiguous auditory filters, thresholds decreased compared to the condition with the spectral energy in only one auditory filter, showing an ability to combine information from the auditory filters and improve the performance.

C. Discussion

The obtained data, in particular those for the ITD conditions, indicate a specific effect of having the signal in a number of contiguous auditory filters. The ability to discriminate between the spatial configurations of the HTC and the BPN increases with increasing signal bandwidth, which may be caused by the peakedness of the temporal envelopes of the presented stimuli. With increasing bandwidth, the temporal envelope of the BPN does not change, except for a relative increase in higher envelope frequencies. The temporal envelope of the HTC, which is similar to the temporal envelope of the BPN when the bandwidth is small, becomes increasingly peaked with an increase in bandwidth, and, therefore, progressively different from the temporal envelope of the BPN.

Figure 6 shows the output for an auditory filter, centered at 550 Hz, computed using a gammachirp filterbank (Irino and Patterson, 2006), for the HTC (left column), the BPN (middle column), and the composite signal (right column). Shown are the results for signal bandwidths of 1 ERB (top row), 3 ERB in noncontiguous auditory filters (middle row), and 3 ERB in contiguous auditory filters (bottom row). For each signal, the insets indicate the peak or crest factor, i.e., the signal's peak amplitude divided by its rms value. The peak factor of the BPN is about the same for all three spectral conditions. The peak factor of the HTC increases with the presence of signal energy in adjacent auditory filters (compare top and bottom panels), but is not affected by adding components at remote spectral regions (compare top and middle panels). A similar change in peak factor is observed for the composite signal. The availability of these distinguishable peaks in the composite signal's envelope apparently enhances the ability to process binaural cues of the signal components at these moments in time, similar to an onset response that enables the processing of signal components at a specific spectrotemporal position to be dominant for a brief period of time (Bregman *et al.*, 1994).

In the next experiment, further evidence is provided for the idea that the ability to discriminate between the spatial configurations of two signals is influenced by their temporal envelope structures. From the previous discussion, this ability may be expected to break down when their temporal envelope structures are more similar. To explore a possible breakdown of discrimination ability, the experiments were partially repeated for changed temporal envelope structures of either the HTC or the BPN.

VII. EXPERIMENT 4

This experiment investigated the effect of temporal envelope structures on the ability to discriminate between the spatial configurations of two spectrally and temporally overlapping signals. Various temporal envelope structure manipulations were applied to either the HTC or to the BPN. The experiment was limited to measuring thresholds for the wideband signal conditions, because these measurements led to the lowest thresholds in the previous experiments.

A. Stimuli

The temporal envelope structure of the HTC was manipulated in three different ways to yield a temporal envelope more similar to that of the BPN. First, a random starting phase was applied to each of the HTC components, which results in a temporal envelope with a low peak or crest factor. The HTC was defined as

$$x(n) = \sum_{k=i}^{j} A \sin\left(2\pi k f_0 \frac{n}{f_s} + \Phi_r\right), \tag{2}$$

where Φ_r is a random number uniformly distributed in the range $[0, 2\pi)$.

Second, both positive and negative Schroeder phases were applied to the HTC components, i.e., a downward and upward frequency sweep of the components, resulting in a flat temporal envelope of the HTC (Schroeder, 1970). The HTC was now defined as

$$x(n) = \sum_{k=i}^{j} A \sin\left(2\pi k f_0 \frac{n}{f_s} + \Phi_S(k-i)\right), \tag{3}$$

with

$$\Phi_S(m) = \frac{\frac{1}{2}\pi + c\pi(m+1)m}{j-i+1}, \tag{4}$$

where $c=1$ for a positive Schroeder phase and $c=-1$ for a negative Schroeder phase. Due to the linear frequency modulation, the excitation peak in the different auditory filters was not synchronized as it was for the zero-starting-phase signal.

Figure 7 displays the waveforms of the four phase conditions as applied to the wideband HTC. Each phase condition results in a different temporal envelope structure. The first panel shows the zero-phase HTC as used in Exp. 1, with its regularly peaked pattern. The second panel shows that the periodic pattern of strong peaks for the zero-phase HTC completely disappears when a random phase is given to each component. The third and fourth panels show the positive and negative Schroeder-phase HTCs, respectively, for which the application of the Schroeder phases to the components results in a periodic signal with a flat temporal envelope.

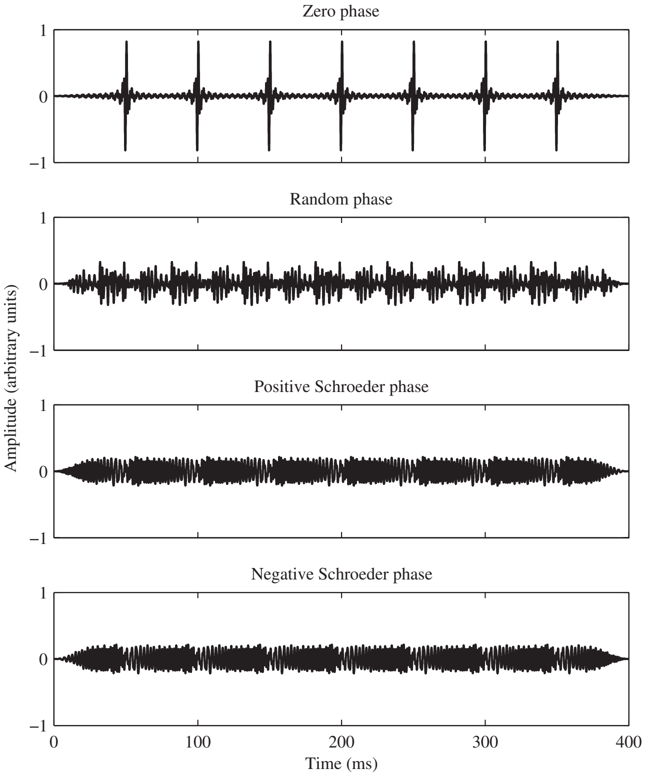

FIG. 7. Four phase conditions of the wideband HTC, resulting in different temporal envelopes. The phase of the components was manipulated to yield temporal envelopes of the HTC more similar to the temporal envelope of the BPN.

Third, the component spacing of the HTC was increased to 40 and 80 Hz, which results in an increase in the number of peaks in the temporal envelope. This manipulation, however, reduces the peak factor after filtering on the basilar membrane, because fewer components fall within the bandwidth of a single auditory filter. The HTC was defined as in Eq. (1), for $f_0=40$ Hz and $f_0=80$ Hz. To preserve the harmonicity of the HTC at multiples of 40 and 80 Hz, the spectral features of both signals had to be slightly changed and were set to a 560 Hz center frequency with a 600 Hz bandwidth for $f_0=40$ Hz, and a 600 Hz center frequency with a 640 Hz bandwidth for $f_0=80$ Hz.

The temporal envelope of the BPN was manipulated by applying a 20 Hz sinusoidal amplitude modulation to yield a temporal envelope with regular peaks at the same period as those of the zero-phase HTC. The amplitude-modulated BPN was defined as

TABLE I. Conditions of Exp. 4: Temporal envelope structure manipulations for the harmonic tone complex (HTC) or the bandpass noise (BPN), including their spectral properties. The manipulation was only applied to the temporal envelope of the signal mentioned; the temporal envelope of the other signal remained the same as in Exp. 1.

Signal	Temporal envelope	CF	BW
HTC	Random phase	550	600
HTC	Positive Schroeder phase	550	600
HTC	Negative Schroeder phase	550	600
HTC	40 Hz fundamental	*560*	600
HTC	80 Hz fundamental	*600*	*640*
BPN	Amplitude modulation (ϕ=0)	550	600
BPN	Amplitude modulation (ϕ=π)	550	600

$$y(n) = \sqrt{\frac{2}{3}}N(n)\left(1+\cos\left(2\pi f_0\frac{n}{f_s}+\Phi_{\mathrm{AM}}\right)\right), \qquad (5)$$

where N is the BPN, and the multiplication with $\sqrt{\frac{2}{3}}$ is applied to preserve the energy of the noise. In combination with the amplitude-modulated BPN, the HTC was presented with a zero starting phase and a 20 Hz fundamental frequency. The HTC and amplitude-modulated BPN were presented either temporally in phase ($\Phi_{\mathrm{AM}}=0$), such that the envelope maxima of both signals coincided, or out of phase ($\Phi_{\mathrm{AM}}=\pi$), such that the envelope maxima of one signal coincided with the envelope minima of the other signal. Table I gives an overview of the signals' temporal envelope structure conditions, including their spectral properties.

B. Results

Figure 8 displays the mean thresholds and the standard errors of the mean of the pooled data of the five subjects for the various temporal envelope structure conditions. The top panel shows the data for the ITD conditions and the bottom panel the data for the ILD conditions. The abscissa indicates the eight temporal envelope structures, including the combination of the wideband zero-phase HTC and the unmodulated BPN from Exp. 1 to the left in both panels. The ordinate indicates the size of the thresholds in microseconds (ITD) or decibels (ILD). The symbols represent the mean thresholds for the manipulations of the temporal envelope of the wideband HTC (open markers) and of the BPN (closed markers).

For the condition with a random starting phase on each component of the HTC, discrimination of the spatial configurations based on interaural time differences was not possible. Discrimination based on interaural level differences was seriously degraded compared to the reference zero-phase condition, and the threshold ILD of 29 dB was beyond the value needed for maximal lateralization.

When positive and negative Schroeder phases were applied to the HTC components, discrimination based on interaural time differences was seriously degraded compared to the 29 μs threshold of the zero-phase condition, with thresholds of 132 and 86 μs, respectively. Discrimination based on interaural level differences was only slightly degraded compared to the 1 dB threshold of the zero-phase condition, with thresholds 1.2 and 0.6 dB higher for the positive and negative Schroeder-phase conditions.

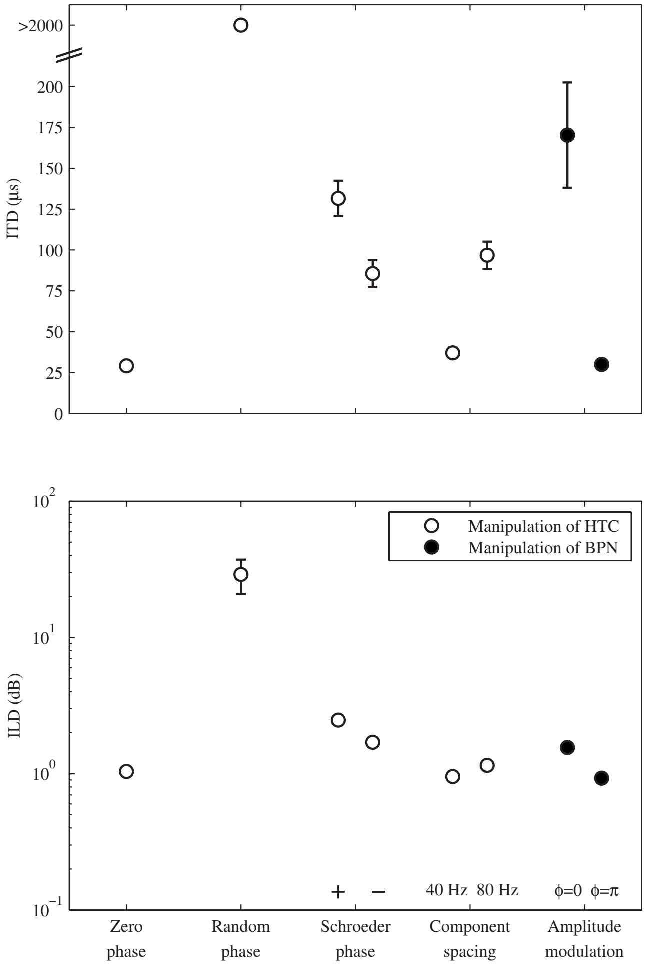

FIG. 8. Mean ITD (top panel) and ILD (bottom panel) thresholds for the temporal envelope structure conditions from Exp. 4. Included for reference is the result for the wideband zero-phase condition from Exp. 1. Error bars represent the standard errors of the mean. Error bars smaller than the symbol size are omitted. Please remember that the total interaural difference cue between the two signals is, in fact, twice the size of these interaural differences.

For the condition with 40 Hz component spacing of the HTC, discrimination based on interaural time differences was, with a threshold of 37 μs, similar to the performance for the reference 20 Hz component spacing condition. For the condition with 80 Hz component spacing, with a threshold of 97 μs, discrimination was seriously degraded. For these two component spacing conditions, discrimination based on interaural level differences was equal to the performance for the zero-phase condition, with thresholds 0.0 and 0.2 dB higher for the 40 and 80 Hz component spacing conditions, respectively. Thus, halving the time between the regular maxima in the temporal envelope of the HTC to 25 ms caused only little change in binaural sensitivity. By again halving the time between regular maxima to 12.5 ms, binaural processing of interaural time differences was made much more difficult. Again, no change in binaural sensitivity to interaural level differences was found, indicating that small opposite interaural level differences for each signal, combined with different temporal envelopes, sufficed to allow discrimination between the spatial configurations of the two signals.

For the condition with the HTC and the amplitude-modulated BPN presented in phase ($\phi_{AM}=0$), discrimination between the spatial configurations of the two signals based on interaural time differences was heavily degraded, with a threshold of 170 μs. As can be seen in Fig. 8, top panel, the performance in this condition differed substantially across subjects. For the condition with the two signals presented out of phase ($\phi_{AM}=\pi$), discrimination ability was equal to the reference zero-phase condition, with a threshold only 1 μs higher. For these two conditions, discrimination based on interaural level differences was similar to the performance for the zero-phase condition, with thresholds 0.6 dB higher and 0.1 dB lower for the in-phase and out-of-phase conditions, respectively. Thus, when the maxima of both signals coincided temporally, the monaural envelopes were apparently more similar, resulting in a substantially higher threshold for discrimination based on interaural time differences. Small level differences between the signals provided sufficient monaural or binaural information to discriminate between their spatial configurations. When the maxima of the signals alternated temporally, the binaural processing of these signals was comparable to that for the reference zero-phase condition, with comparable performance in discrimination between the spatial configurations of the two signals.

C. Discussion

Because the long-term interaural cross-correlation patterns or long-term patterns after equalization-cancellation for the composite signals in the ITD conditions were indistinguishable between the target and the reference intervals (see Fig. 2), short-term binaural processing must have played a role in discriminating between the spatial configurations of the two signals. One interpretation is that in order to achieve the observed lateralization of the individual signals', the binaural system would require sufficient information about their temporal envelopes to distinguish between them and assign the spatial cues to the correct signal. Then, the monaural temporal envelope cues somehow would have to support the selection of temporal intervals at which the one or the other signal was dominant, and facilitate the organization of temporally varying interaural time differences within the period of the HTC.

In line with this view, it follows that for discrimination between the spatial configurations of two spectrally and temporally overlapping signals, the temporal envelopes of the signals need to be sufficiently different, either in terms of the degree of modulation or the relative timing of the envelope maxima, in order to be able to link the pattern of temporally changing interaural time differences to the envelope maxima of the signals. For the selection of interaural time differences within a composite signal, these interaural differences could be emphasized during the periods when a single source is dominant in one or more auditory filters.

A recent approach to selecting these time instants by analyzing the interaural coherence is the directional cue selection model of Faller and Merimaa (2004). In this model, the time instants at which a single sound is dominating in an acoustic scene are identified as moments where the interaural coherence is high. Here, a perceptually-motivated adaptation of their model, using a gammachirp filterbank (Irino and Patterson, 2006) instead of a gammatone filterbank and peripheral compression stages, is used to determine the short-term cross-correlation patterns for the stimuli used in these experiments.

Figure 9 shows the results for the wideband condition from Exp. 1, in which the HTC had its components in zero starting phase, and the noise signal was unmodulated. The three top panels show the individual signals and the composite signal for one ear, while the fourth panel shows the output of a single auditory filter centered at 550 Hz from the gammachirp filterbank, followed by a hair cell model. The fifth panel shows the cross-correlation pattern between the left- and right-ear signals, computed with an exponentially decaying window with a time constant of 10 ms (Faller and Merimaa, 2004). High correlation values are indicated by darker colors in the figure, and the maximum of the cross-correlation function at each temporal instant is traced by the white curve. The absolute interaural time difference of the signals in this and the following examples was 100 μs, a value much larger than the experimental thresholds observed. The maximum of the cross-correlation pattern varies periodically between the values of ±100 μs (indicated by the dashed lines) in close synchrony with the monaural envelope of the composite signal, enabling the determination of the individual signal components' lateralizations.

When analyzing the conditions in which ITD discrimination thresholds were successfully obtained, i.e., the contiguous bandwidth conditions of Exp. 3, the various component spacings of the HTC, and the amplitude-modulated BPN of the current experiment, similar observations are made. In all these analyses, the cross-correlation pattern is found to follow the monaural envelope of the composite signal in close synchrony. For example, Fig. 10 shows the temporal envelopes and interaural cross-correlation pattern for the zero-phase HTC and the amplitude-modulated BPN presented with their envelope maxima temporally out of phase.

In contrast to these conditions, Fig. 11 shows the corresponding patterns for the combination of a random-phase HTC and the unmodulated BPN. Changing the temporal envelope structure of the HTC reduced its peak factor and made it similar to the temporal envelope of the BPN, instead of regularly peaked as in the zero-phase condition. Changing the temporal envelope structure also influenced the regular cross-correlation pattern of interaural differences. Here, neither the monaural envelope nor the cross-correlation pattern allow determination of time instants at which one of the two constituent signals is dominant. As a consequence, no information is available for linking specific moments in the internal representation of interaural delay to those of the monaural temporal envelopes, and thus to identify the target and reference intervals based on the available spatial information.

Similar observations can be made for the conditions in which ITD discrimination thresholds could not be measured,

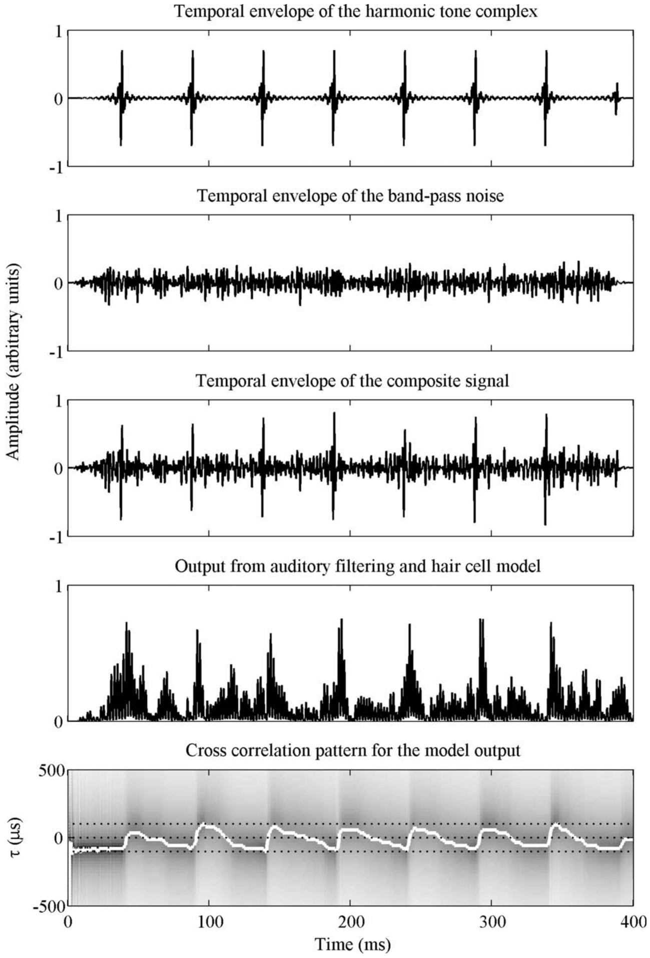

FIG. 9. The relation between interaural cross-correlation pattern and the temporal envelope, for the wideband condition with the components of the HTC in zero phase from Exp. 1. The three top panels show the temporal envelopes of the individual wideband signals and their composite signal. The HTC and the BPN in this example have 100 μs interaural time differences with opposite signs. The composite left- and right-ear signals were used as input to the directional cue selection model of Faller and Merimaa (2004). The two lower panels show the results from the model, i.e., the output of the auditory filter centered at 550 Hz and hair cell model, and the interaural cross-correlation pattern computed from this output, respectively. The dashed lines in the lower panel represent the 0 and ±100 μs interaural time differences.

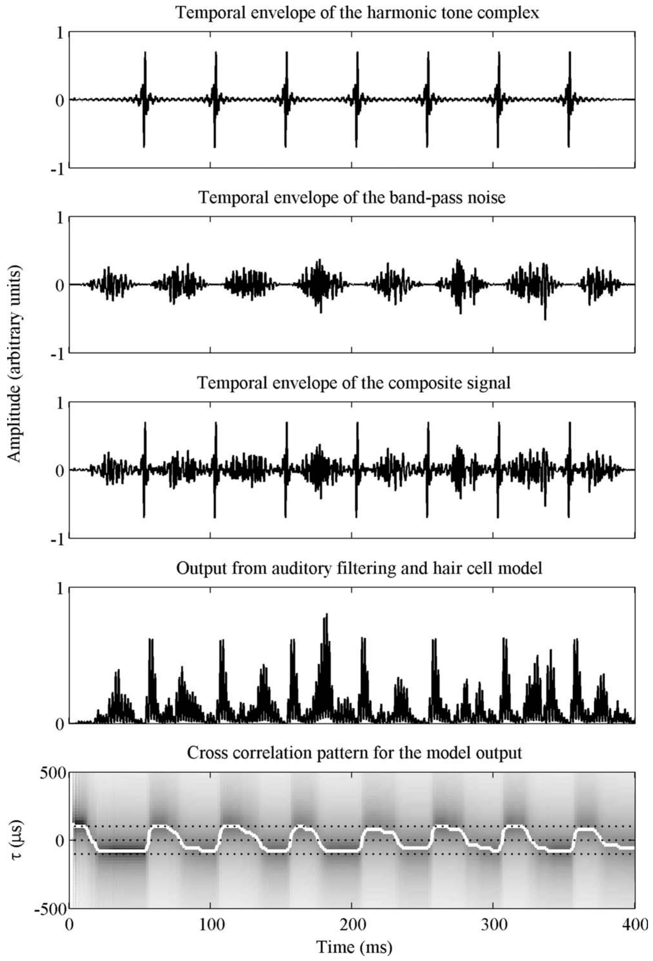

FIG. 10. The relation between interaural cross-correlation pattern and the temporal envelope, for the condition with the zero-phase HTC and the amplitude-modulated BPN presented with their envelope maxima temporally out of phase. Similar to Fig. 9, the panels show the temporal envelopes of the individual signals, their composite signal, the output of the auditory filter and hair cell model, and the interaural cross-correlation pattern of this output. The dashed lines in the lower panel represent the 0 and ±100 μs interaural time differences.

as, for example, the narrowband condition of 1 ERB centered at 550 Hz from Exp. 1, as shown in Fig. 12. In this narrowband condition, the peaks of the HTC, although regularly spaced in time, are insufficiently peaked to dominate the temporal envelope of the composite signal in the same manner as seen in the wideband condition. As in the random-phase condition, the information from the internal representation of the interaural delay could not be linked to the information from the monaural temporal envelopes.

From the regular short-term interaural cross-correlation pattern, as displayed in the bottom panel of Figs. 9 and 10, it may be argued that identification of the target and reference intervals could have been established by distinguishing between the regularly changing pattern of the cross-correlation function in the target interval and its interaural inverse in the reference intervals. The pattern of dynamically changing interaural time differences itself may then have provided sufficient information to the binaural system to perform the discrimination task. For the random-phase and narrowband conditions, in which no threshold could be obtained, the cross-correlation pattern does not exhibit a systematic asymmetrical pattern, as can be seen in Figs. 11 and 12. Using the asymmetrical cross-correlation patterns that are mirrored for target and reference intervals would require subjects to be able to track fast chances in binaural cues, such as observed in the fifth panel synchronous to the peaks of the auditory filter output in the fourth panel. This requirement, however, contrasts to earlier findings on the inability to track fast changes in binaural cues (Grantham and Wightman, 1978). In combination with the ability to lateralize the individual signals, as reported by the subjects (see discussion of Exp. 1), it is considered unlikely that subjects only used the interaural cross-correlation pattern to identify target and reference intervals. In addition, when the same analysis was applied to the condition with the zero-phase HTC and the amplitude-modulated BPN presented with their envelope maxima temporally in phase, see Fig. 13, the cross-correlation pattern

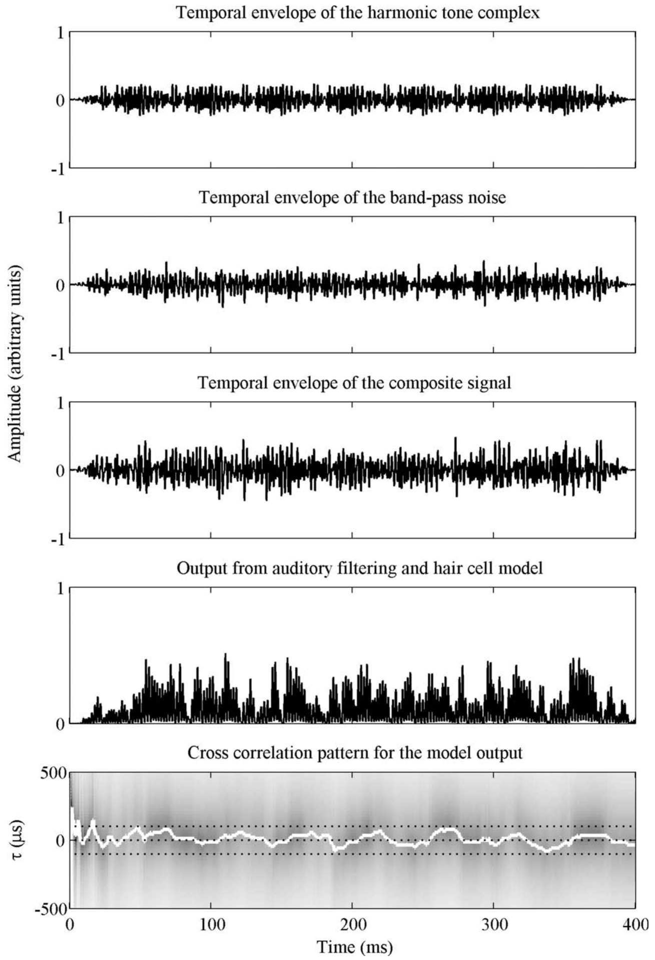

FIG. 11. The relation between interaural cross-correlation pattern and the temporal envelope, for the condition with the components of the HTC in random phase. Again, the panels show the temporal envelopes of the individual signals, their composite signal, the output of the auditory filter and hair cell model, and the interaural cross-correlation pattern of this output. The dashed lines in the lower panel represent the 0 and ±100 μs interaural time differences.

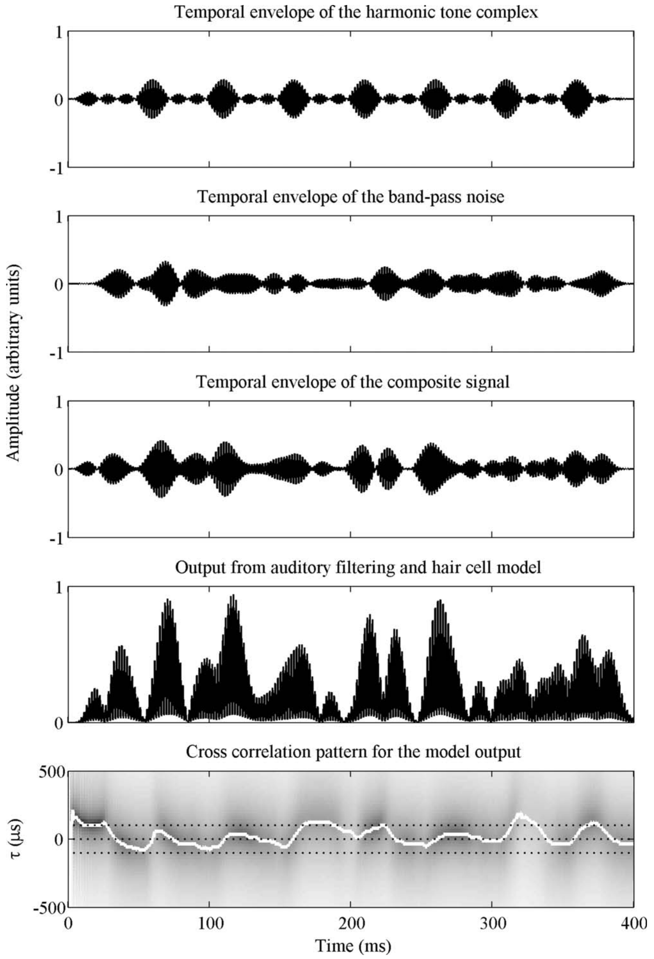

FIG. 12. The relation between interaural cross-correlation pattern and the temporal envelope, for the narrowband condition of 1 ERB centered at 550 Hz from Exp. 1. The panels show the temporal envelopes of the individual signals, their composite signal, the output of the auditory filter and hair cell model, and the interaural cross-correlation pattern of this output. The dashed lines in the lower panel represent the 0 and ±100 μs interaural time differences.

exhibits similar fast changes in the positions of the maxima as displayed in Figs. 9 and 10. If such fast changes, with different orientations for the target and reference intervals, were the dominant cue for identification of these intervals, it is hard to understand why patterns as in Figs. 9 and 10 yield thresholds of about 30 μs, while a pattern as in Fig. 13 the threshold was increased to about 170 μs. From these considerations, it is concluded that synchrony between the temporal envelopes and the pattern of changing interaural time differences seems to be necessary, but not sufficient, to identify target and reference intervals.

The principle of determining a time instants' dominant signal and processing the momentary spatial cues would also explain the increased difficulty to discriminate between the spatial configurations for several ITD conditions in the current experiment. For the Schroeder-phase conditions, the effect of using Schroeder phases for the HTC is twofold. Due to the frequency-dependent group delay, maxima in the temporal envelope are asynchronous across frequency. Therefore, no single moment in time can be defined where the HTC dominates across all frequencies synchronously. Furthermore, positive Schroeder phases compensate for the phase dispersion on the basilar membrane, resulting in a larger peak factor at the output of the basilar membrane, while negative Schroeder phases result in a lower peak factor (Kohlrausch and Sander, 1995). Apparently, synchrony across frequency is not essential, although it influences the ability to discriminate between the target and reference intervals. For the 80 Hz component spacing condition, the reduced temporal interval between peaks may have reduced the peak factor of the temporal envelopes after filtering on the basilar membrane such that dominant signal components were more difficult to be distinguished monaurally. In addition, here the temporal resolution of the binaural system may have been insufficient to accurately follow the faster switching between the opposite binaural cues. For the HTC presented in phase with the amplitude-modulated BPN, the determination of the dominant signal at a certain time instant may have been inhibited by the similarity of the signals' temporal envelopes and the temporal coincidence of both envelope maxima and minima.

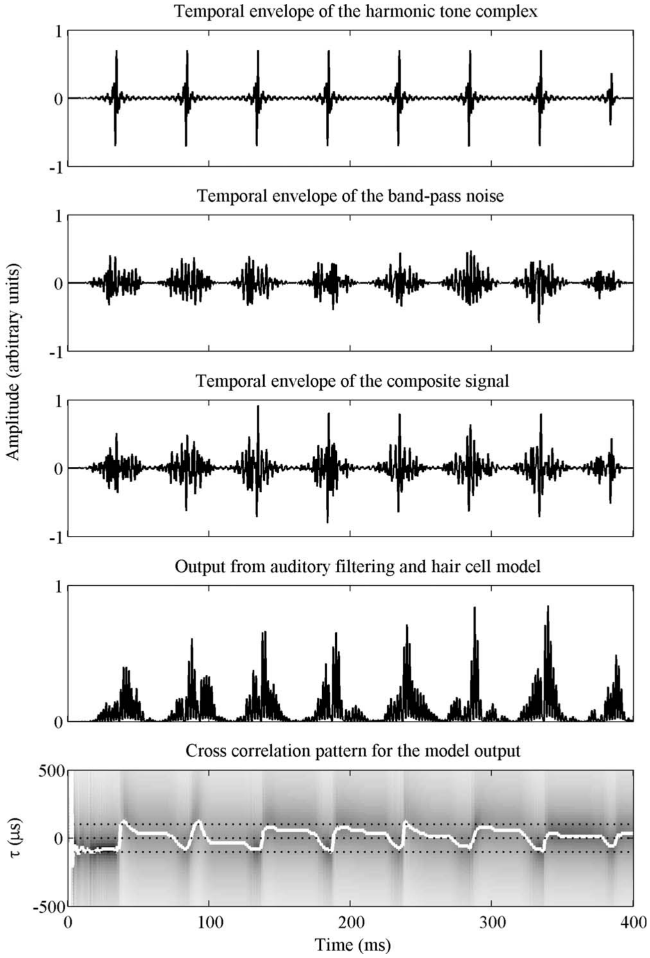

FIG. 13. The relation between interaural cross-correlation pattern and the temporal envelope, for the zero-phase HTC and the amplitude-modulated BPN with their envelope maxima temporally in phase. The panels show the temporal envelopes of the individual signals, their composite signal, the output of the auditory filter and hair cell model, and the interaural cross-correlation pattern of this output. The dashed lines in the lower panel represent the 0 and ±100 μs interaural time differences. In contrast to the conditions shown in Figs. 9 and 10, subjects had more difficulty to identify target and reference intervals in this condition.

VIII. GENERAL DISCUSSION

In the current study, the ability to segregate two spectrally and temporally overlapping signals based on differences in temporal envelope structure and binaural cues was investigated. As an indication for segregation ability, threshold interaural time and level differences for discriminating between the signals' spatial configurations were measured for various signal bandwidths, center frequencies, and temporal envelopes. The interpretation of these measures is only valid if the target and reference intervals of the three-alternative forced-choice experiment cannot be identified otherwise than by segregating and lateralizing the constituents of the composite signal in each interval. It was shown that the application of a rove on the interaural time difference for the composite signal did not degrade the discrimination of the signals' spatial configurations. It seems unlikely that the results can be explained by the subjects' perception of a change of the overall lateralizations of the composite signal between the target and reference intervals. Alternatively, the identification of the intervals by distinguishing between their mirrored patterns in the short-term cross-correlation functions may explain why for narrowband signals and the random-phase HTC discrimination based on interaural time differences was not possible. The corresponding dynamic patterns of interaural time differences have no systematic asymmetry that would allow discrimination between the intervals. However, it was shown that similar asymmetrical fast changes in the maxima of the cross-correlation function can be found for the ITD conditions that yielded the lowest and the highest threshold in discrimination between the target and reference intervals. In addition, the use of only the mirrored cross-correlation patterns does not match the subjects' report on the use of the lateral position of the HTC for identifying target and reference intervals.

For the ITD conditions, the spatial configurations of the signals could only be discriminated when the spectral energy was in multiple contiguous auditory filters, with an increase in performance with an increase in bandwidth. This bandwidth dependency may be related to across-frequency integration of the coherence between temporal envelopes in the stimulated auditory filters (cf. Trahiotis and Stern, 1994). However, it is unclear how the across-frequency integration could account for the inability to discriminate between the spatial configurations of signals that had their energy in noncontiguous bands. This inability may be attributed to the observed difference in peak factor of contiguous-band versus noncontiguous-band signals (see Fig. 6). For the contiguous-band HTC, the signal components in adjacent auditory filters contributed to the within-channel peak factor, which did not occur for the noncontiguous-band HTC and the noncontiguous-band and contiguous-band BPNs. A similar contribution of signal components in adjacent auditory filters to within-channel perception has also been shown in the context of comodulation masking release (Verhey *et al.*, 1999).

The obtained results show that the differences in temporal envelope structures of the two signals influence the discriminability between their spatial configurations. The results are consistent with the hypothesis that the analysis of the monaural information from the temporal envelopes at each ear supports the attribution of binaural information to signal components. For the ITD conditions, discrimination of the signals' spatial configurations was found to be inhibited for conditions in which their temporal envelopes were similar or the maxima and minima of the temporal envelopes coincided. For instance, for the narrowband or noncontiguous-band signal conditions, the temporal envelopes of the two signals were much more similar and discrimination was not possible. Also, for the wideband random-phase HTC and the in-phase amplitude-modulated BPN conditions, it was more difficult to distinguish between the temporal envelopes of the two signals, resulting in either the complete inability or the increased difficulty to discriminate between the signals' spatial configurations. For the contiguous wideband signal conditions, the HTC had a considerably higher peak factor than the BPN and discrimination was possible. This reasoning may explain the lack of effect from interaural time differences as a grouping cue in the experiments of Culling and Summerfield (1995). By using various combinations of filtered bands of continuous noise, their vowel-like stimuli may have been composed in such a way that the temporal envelopes of the constituent signals were too similar to facilitate attribution of the binaural information to either of them.

For the ILD conditions, the signals' spatial configurations could be discriminated within a single auditory filter, as long as the phase spectrum of the HTC was not random. The left and right ears received different relative levels of the HTC and the BPN. Therefore, unlike the ITD conditions, the envelope was more peaked in the ear where the HTC was louder. Listening to only the left or right ear, in principle, provided sufficient cues to discriminate between the signals' spatial configurations. For the narrowband conditions, the performance based on monaural listening was even better than the one based on binaural listening. This finding suggests that the processing of binaural information from each left-right pair of auditory filters interfers with the processing of monaural information from the individual auditory filters. For the wideband conditions, performance based only on either monaural or binaural listening was found to be very similar. Therefore, it is not possible to distinguish between monaural or binaural processing on the basis of these results.

In general, differences in the temporal envelope of signals seem to improve the use of their binaural cues for segregation. These differences even may have allowed monaural segregation of the composite signal's components by perceiving continuity in temporal envelope features that are "glimpsed" at momentary differences in level (cf. Noble and Perrett, 2002; Shinn-Cunningham, 2005; Cooke, 2006). To establish such segregation, the auditory system may adopt simultaneously several hypotheses about how the monaural input signals can be segregated based on monaural temporal envelope features. The binaural cues corresponding to each of the envelope features would then assist in selecting the most likely hypothesis, based on the assumption that within one auditory stream the binaural cues have to be coherent across time. Also some top-down processing may be assumed in this context, as prior knowledge about a certain signal would enhance its identification. Thus, for the selection of binaural cues within composite signals, interaural time and level differences are best considered only at time instants at which the sound of a single source is dominant in one or more auditory filters. In the directional cue selection model of Faller and Merimaa (2004), these time instants are identified by analyzing the interaural coherence, and correspond well to the moments in time at which glimpsing would be operative.

In a previous study (van de Par *et al.*, 2005), the discrimination between the same HTC and BPN as used in the current study was investigated, when both were presented interaurally out of phase within an in-phase noise masker. The results showed an ability to discriminate between the two signals at signal-to-masker levels well below monaural detection thresholds. This finding suggests that within the binaural display, i.e., the internal representation of binaural cues, information is available about the temporal structure of the out-of-phase signal. An equalization-cancellation stage (Durlach, 1963; Breebaart *et al.*, 2001) could, in principle, provide such information, because it would cancel the noise masker and not the out-of-phase signal. It would require, however, that within the binaural output, the capacity to process temporal information is sufficiently good to distinguish between the signals.

In the current study, an equalization-cancellation stage could remove one of the two signals in a similar way, allowing for the temporal processing of the other signal. However, the sharp increase in threshold (from 29 via 37 to 97 μs) from decreasing the period of the HTC (from 50 via 25 to 12.5 ms) indicates the vicinity of an upper limit for the temporal resolution at the output of a binaural display for the current experiments. From the current results, it is not evident how equalization-cancellation processing could explain that subjects had such difficulty to identify the target and reference intervals in the condition with the zero-phase HTC and the amplitude-modulated BPN presented in phase.

The glimpsing of signal properties as described above could provide an alternative explanation for the reduced ability to discriminate between the spatial configurations of the zero-phase HTC and the amplitude-modulated BPN presented in phase. For this condition, the peaks in the envelopes of both signals coincide, and the level difference between the maxima of the signals is smaller than for the nonmodulated condition due to maintaining the same overall level for the amplitude-modulated BPN. This made the temporal envelopes similar, and may have facilitated perceptual merging of the two signals into one auditory object. Because of the "discontinuity" of the BPN in between the regular peaks of the composite signal, there was simply less evidence for the presence of a second auditory object with a different lateralization. Similarly, for the random-phase condition there is no evidence in the monaural envelope of the composite signal that allows glimpsing of the constituent signals. This explanation supports the idea that monaural analysis of temporal envelopes on perceived continuity of the glimpsed signal components is required before binaural processing can take place.

IX. CONCLUSION

The ability to segregate two spectrally and temporally overlapping signals based on differences in temporal envelope structure and binaural cues was investigated. As an indication for segregation ability, threshold interaural time and level differences were measured for discrimination between the spatial configurations of the two signals. Discrimination based on interaural level differences was good for all conditions, although absolute thresholds depended on signal bandwidth and center frequency. Discrimination based on interaural time differences depended on the signals' temporal envelope structures.

The HTC and BPN were presented with interaural differences of the same absolute value, but with opposite signs, to yield lateralization to different sides of the medial plane. This way, the composite signal's long-term interaural cross-correlation patterns or the long-term patterns after equalization-cancellation were indistinguishable between the three intervals of the forced-choice procedure, and could not facilitate identification of target and reference intervals. For successful identification of the intervals in the ITD conditions, the binaural system must have been capable of processing changes in interaural time differences within the period of the HTC. Such processing would require short-term analysis of the binaural cues present in the composite signal, and association of the specific pattern of time-varying binaural cues with the target and reference intervals.

The reported lateralization of the individual signals and the obtained experimental results support the idea that the binaural system uses the short-term monaural information, that is glimpsed from the temporal envelopes at each ear, to process the binaural information of the underlying signal component. This processing facilitates segregation and lateralization of a composite signal's constituent elements. The current findings suggest that monaural information from the temporal envelopes influences the use of binaural information in the perceptual organization of signal components.

ACKNOWLEDGMENTS

We would like to thank Barbara Shinn-Cunningham, Aki Härmä, Nicolas LeGoff, the associate editor Ruth Litovsky, Steve Colburn, and two anonymous reviewers for their valuable comments on the earlier versions of this paper.

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

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

Bregman, A. S., Ahad, P., Kim, J., and Melnerich, L. (**1994**). "Resetting the pitch-analysis system: 1. Effects of rise times of tones in noise backgrounds or harmonics in a complex tone," Percept. Psychophys. **56**, 155–162.

Buell, T. N., and Hafter, E. R. (**1991**). "Combination of binaural information across frequency bands," J. Acoust. Soc. Am. **90**, 1894–1900.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Culling, J. F., and Summerfield, Q. (**1995**). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am. **98**, 785–797.

Darwin, C. J., and Hukin, R. W. (**1997**). "Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity," J. Acoust. Soc. Am. **102**, 2316–2324.

Darwin, C. J., and Hukin, R. W. (**1998**). "Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony," J. Acoust. Soc. Am. **103**, 1080–1084.

Darwin, C. J., and Hukin, R. W. (**1999**). "Auditory objects of attention: The role of interaural time differences," J. Exp. Psychol. Hum. Percept. Perform. **25**, 617–629.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

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.

Gallun, F. J., Mason, C. R., and Kidd, G., Jr., (**2007**). "The ability to listen with independent ears," J. Acoust. Soc. Am. **122**, 2814–2825.

Grantham, D. W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511–523.

Heller, L. M., and Trahiotis, C. (**1995**). "The discrimination of samples of noise in monotic, diotic, and dichotic conditions," J. Acoust. Soc. Am. **97**, 3775–3781.

Hukin, R. W., and Darwin, C. J. (**1995**). "Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel," J. Acoust. Soc. Am. **98**, 1380–1387.

Irino, T., and Patterson, R. D. (**2006**). "A dynamic compressive gammachirp auditory filterbank," IEEE Trans. Audio, Speech, Lang. Process. **14**, 2222–2232.

Kohlrausch, A., and Sander, A. (**1995**). "Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets," J. Acoust. Soc. Am. **97**, 1817–1829.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Noble, W., and Perrett, S. (**2002**). "Hearing speech against spatially separate competing speech versus competing noise," Percept. Psychophys. **64**, 1325–1336.

Schroeder, M. R. (**1970**). "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation," IEEE Trans. Inf. Theory **16**, 85–89.

Shackleton, T. M., and Meddis, R. (**1992**). "The role of interaural time difference and fundamental frequency difference in the identification of concurrent vowel pairs," J. Acoust. Soc. Am. **91**, 3579–3581.

Shinn-Cunningham, B. G. (**2005**). "Influences of spatial cues on grouping and understanding sound," Proceedings of Forum Acusticum, pp. 1539–1544.

Stellmack, M. A., and Lutfi, R. A. (**1996**). "Observer weighting of concurrent binaural information," J. Acoust. Soc. Am. **99**, 579–587.

Stern, R. M., Trahiotis, C., and Ripepi, A. M. (**2006**). "Fluctuations in amplitude and frequency enable interaural delays to foster the identification of speech-like stimuli," in *Dynamics of Speech Production and Perception*, edited by P. Divenyi, S. Greenberg, and G. Meyer (IOS, Amsterdam), pp. 143–151.

Trahiotis, C., and Stern, R. M. (**1994**). "Across-frequency interaction in lateralization of complex binaural stimuli," J. Acoust. Soc. Am. **96**, 3804–3806.

van de Par, S., Kohlrausch, A., Breebaart, J., and McKinney, M. (**2005**). "Discrimination of different temporal envelope structures of diotic and dichotic target signals within diotic wide-band noise," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 398–404.

Verhey, J. L., Dau, T., and Kollmeier, B. (**1999**). "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model," J. Acoust. Soc. Am. **106**, 2733–2745.