Source segregation based on temporal envelope structure and binaural cues

Steven van de Par[1], Othmar Schimmel[2], Armin Kohlrausch[1,2] and Jeroen Breebaart[1]

[1] Digital Signal Processing Group, Philips Research Laboratories, Eindhoven, {Steven.van.de.Par, Armin.Kohlrausch, Jeroen.Breebaart}@philips.com
[2] Human-Technology Interaction, Technische Universiteit Eindhoven, O.V.Schimmel@tm.tue.nl

1 Introduction

The lateralization of a single auditory object is mediated by interaural time delays (ITDs) and interaural level differences (ILDs). In daily life, listeners regularly encounter multiple auditory objects simultaneously, and it is of interest to learn to what extent and how listeners can localize each object.

When the spectra of the objects differ sufficiently, the binaural cues within each pair of auditory filters would result from the auditory object that has most energy in the frequency range of that auditory filter. Although in principle this could provide cues for the separation of auditory objects, in experiments using simultaneously presented noises shaped to represent different vowels, subjects could not use ITD cues for segregation (Culling and Summerfield 1995).

In this paper we want to investigate whether listeners are able to discern the lateralization of two simultaneously presented auditory objects with different temporal structures that are spectrally fully overlapping. Both objects, (band-pass noise and a harmonic tone complex) resulted in nearly the same spectral excitation pattern, while due to their different temporal structures, the objects could be well distinguished when listened to in isolation. By presenting these two auditory objects simultaneously and with different binaural properties, binaural cues stemming from both objects are equally reflected within each auditory filter. Therefore, in order to correctly lateralize one of the two objects, listeners would need to somehow separate the binaural cues within each single auditory filter and couple these cues to the auditory objects.

Headphone lateralization experiments were done for different bandwidths and center frequencies in a similar way as in our earlier ISH contribution that was dealing with the discrimination of the same two auditory objects based on binaural cues (van de Par, Kohlrausch, Breebaart and McKinney 2005).

2 Experiment I

This experiment was carried out to investigate the ability of listeners to correctly lateralize two simultaneously presented, spectrally fully overlapping signals with opposing binaural cues. In order to perform this lateralization, listeners must be able to segregate the signals based on their opposing binaural (ITD or ILD) cues and their different temporal structure. This task will be termed "discrimination task", because it required listeners to discriminate between lateralizations of the signals.

Besides measuring lateralization thresholds for the two signals presented simultaneously, as a reference, also lateralization thresholds were measured for each of the signals separately. These tasks will be termed "detection task", because they deal with basic lateralization detection thresholds.

2.1 Method and stimuli

For the discrimination task, stimuli consisted of two signals: a band-pass noise (BPN) with a flat spectral envelope, and a harmonic tone complex (HTC) with 20-Hz component spacings and a sinusoidal phase spectrum. Both signals had the same spectral range but differed in their temporal envelope structures. The level for each of the two signals was 65 dB SPL. The two reference intervals had binaural cues such that the BPN was lateralized to the right and HTC to the left using identical but opposing binaural cues. In the target interval the lateralizations of both signals were reversed.

In the narrowband conditions, bandwidths were one critical band wide (1 ERB) and centered at 280 Hz, 550 Hz or 800 Hz. In the wideband condition, the bandwidth was 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, and were separated by 300 ms of silence. For the ILD conditions, level changes were such that the total added energy of left and right target signal was always constant.

To compare performance in the discrimination task, threshold ITDs and ILDs were also measured for the BPN and the HTC in isolation. The two reference intervals had binaural cues such that the signal was lateralized to the left. The target interval had identical binaural cues such that the signal was lateralized to the right. Thus, total interaural differences between reference and test intervals were twice the values as reported hereafter.

The method used for measuring lateralization thresholds was a three-interval, three-alternative forced-choice adaptive-tracking procedure. The adaptive variable (ITD or ILD) was adjusted according to a two-down one-up rule, to track the 70.7%-correct response level. The initial adaptive variable was adjusted by multiplying or dividing the adaptive variable 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 value of 1.12 ($=10^{1/20}$) was reached. Another eight reversals were measured at this step size and the median of these eight levels was used as an estimate of threshold. Feedback was provided after each trial.

For each condition, at least four attempts were made by each subject to measure a threshold. However, when the adaptive variable exceeded a certain maximum value, the tracking procedure was terminated and no threshold was measured. For these conditions, measurements were either repeated to obtain a total of at least four proper threshold values, or, for the most difficult conditions, measurements were stopped and thresholds (if any) discarded. Five normal hearing male subjects, including the four authors, participated in the experiments.

2.2 Results and discussion

The thresholds for each condition were pooled, and five severe outliers (thresholds at more than three times the interquartile range) out of 426 thresholds were removed from the data. Figure 1 shows the average detection and discrimination threshold and standard error of the mean for the various conditions.

For the ITD conditions (left panel), detection thresholds for lateralization changes of the narrowband BPN and HTC are about 10-20 μs, while for these narrowband conditions a discrimination threshold could not be obtained. Detection thresholds for wideband BPN and HTC are similar to the lowest of the narrowband detection thresholds, about 12 μs, while the wideband discrimination threshold is about 29 μs. These data indicate that segregation of the BPN and the HTC by interaural time delays is not possible when their spectral energy is limited to one auditory filter.

For the ILD conditions (right panel), all detection thresholds for BPN and HTC are about 1 dB independent of center frequency and bandwidth, while narrowband discrimination thresholds are about 5-32 dB. Note that this 32-dB threshold at the 280-Hz center frequency is beyond the plausible range for localization. For these narrowband conditions, discrimination thresholds decrease with increasing center frequency. The wideband discrimination threshold is nearly equal to its corresponding detection thresholds. These data indicate that, in contrast to ITD

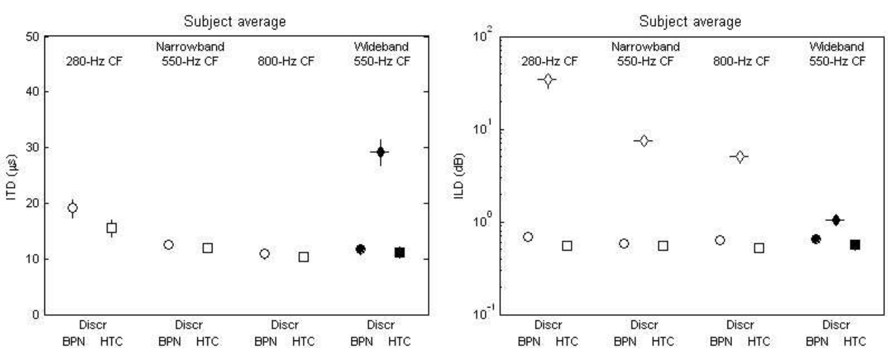

Fig. 1. Average detection thresholds for BPN (circles) and HTC (squares), and discrimination thresholds (diamonds) are shown for ITD (left panel) and ILD (right panel) conditions for narrowband (open markers) conditions and wideband (closed markers) conditions. Vertical lines indicate the standard error of the mean.

narrowband conditions, segregation by interaural level differences of the narrowband noise and the harmonic tone complex is possible within one auditory filter, although thresholds are very high. Analogous to the ITD conditions, segregation is best for conditions where the stimulus spectrum covers multiple auditory filters. From this finding we conclude that segregation by interaural level differences is not only depending on the bandwidth of the signals, but also on the center frequency of the auditory filter.

Overall we can conclude that spectral energy in multiple auditory filters facilitates segregation by binaural listening of spectrally fully overlapping concurrent sound sources.

3 Experiment II

This experiment was defined to investigate to what extent the ILD thresholds for the discrimination task obtained in Exp. I can be understood in terms of listening with one ear only.

3.1 Method and stimuli

The method was the same as in Exp. I, and measurements were limited to discrimination thresholds and to four of the five subjects.

The stimuli were the same as the ILD conditions from Exp. I, only now the right ear signal was presented diotically. In this way, only monaural level cues resulting from the different temporal envelope structures of the HTC and the BPN were available for discriminating between the target and the reference intervals.

3.2 Results and discussion

The thresholds for each condition were pooled, and one severe outlier out of 64 thresholds was removed from the data. Table 2 shows the subject average of monaural and previously established binaural thresholds and standard error of the mean for the various bandwidths and center frequencies.

As for binaural thresholds from Exp. I, monaural thresholds in the narrowband conditions decreased with an increase in center frequency. However, the monaural thresholds were much lower than binaural thresholds. Thresholds for the wideband condition were similar, indicating that monaural and binaural discrimination were equivalent.

	Narrowband			wideband
	280-Hz	550-Hz	800-Hz	550-Hz
Monaural [in dB]	4.8 (0.3)	3.0 (0.2)	2.6 (0.2)	1.2 (0.1)
Binaural [in dB]	34.2 (6.9)	7.4 (0.8)	5.1 (0.8)	1.0 (0.1)

Table 2. Average discrimination thresholds and standard errors of the mean (in brackets) for various bandwidths and center frequencies.

From these findings, we conclude that for narrowband signals binaural stimulus presentation can actually reduce discrimination performance in the case of ILD cues. In other words, in the binaural ILD discrimination condition, much better performance could be obtained if listeners were able to listen with only one ear.

4 Experiment III

This experiment was defined to explore the effect of temporal envelope structure on the ability to segregate two spectrally overlapping signals by binaural cues.

4.1 Method and stimuli

The temporal envelopes of the BPN and HTC were manipulated in two ways. First, a random phase was applied to the HTC components, to yield a temporal envelope more similar to that of the BPN. Second, a 20-Hz amplitude modulation (AM) was applied to the BPN, to yield a temporal envelope more similar to that of the sine-phase HTC. The AM BPN and the HTC were presented temporally in-phase ($\phi=0$), such that the envelope maxima of both signals coincided, or out-of-phase ($\phi=\pi$).

The method was the same as in Exp. I, except that measurements were limited to discrimination thresholds for the wideband condition, because these measurements led to the lowest thresholds. The four subjects of Exp. II participated.

4.2 Results and discussion

The thresholds for each condition were pooled, and no severe outliers were found. Table 3 shows the subject average and standard error of the mean for the different temporal envelope structure conditions (including the sine-phase HTC of Exp. I).

	Sine	Random	$\phi=0$	$\phi=\pi$
ITD [in μs]	29 (2)	-	-	29 (3)
ILD [in dB]	1.0 (0.1)	29.0 (8.2)	1.5 (0.1)	1.0 (0.1)

Table 3. Average discrimination thresholds and standard errors of the mean (in brackets) for various temporal envelope structures of HTC (sine/random phase) and AM BPN ($\phi=0$/$\phi=\pi$).

For random phase conditions, segregation by ITDs was not possible, and segregation by ILDs was seriously degraded. For the conditions with the AM BPN and HTC presented in-phase ($\phi=0$), segregation by ITDs was also not possible, and for ILDs segregation was slightly reduced compared to the sine-phase condition. For conditions with the AM BPN and HTC presented out-of-phase ($\phi=\pi$), segregation was normal for both binaural cues.

From these findings, we conclude that in order to be able to segregate and lateralize the two signals, it is important that the temporal envelope structure of both signals is different, either due to different degrees of modulation or due to a difference in the relative timing of envelope maxima.

5 Discussion and conclusions

This study investigated to what extent two signals with a very similar spectral envelope but different temporal structures can be segregated based on binaural cues. From Exp. I, it appears that listeners can indeed segregate such signals based on a difference in binaural cues of the two signals if the bandwidth of the signals exceeds one critical band.

When we consider the ILD conditions, as was shown in Exp. II, the left or right ear signals by themselves provide monaural cues that are sufficient to distinguish reference and target intervals. Therefore, it is not certain that in Exp. I, the lateralization thresholds for the ILD conditions are necessarily based on binaural processing. Interestingly, for narrowband conditions, the performance based on listening only with the right ear signal was better than based on binaural listening, while in the wideband condition, performance was very similar. This suggests that for narrowband conditions, binaural fusion is mandatory and listening with one of the two ears is not possible.

When we consider the ITD conditions, segregation or lateralization of both signals can only be explained based on binaural processing. However, the binaural cues are mixed equally within each critical band and therefore we have to assume that, somehow, the auditory system is able to separate the cues corresponding to both signals and couple them to the two signals.

It is not clear, however, what would be the nature of binaural processing that allows separation of the ITD cues of both signals in Exps. I & III. It appears that a prerequisite for being able to segregate two binaural signals is that both signals have a distinctly different temporal envelope extending across several auditory filters. For the random-phase HTC and for the in-phase AM BPN, temporal envelopes of both signals were correlated and segregation was impossible. For the sine-phase HTC and the out-of-phase AM BPN, temporal envelopes were much less correlated and lateralization performance was not too dissimilar from that of the signals in isolation.

In addition, it appears that temporal envelopes resulting from narrowband signals by themselves do not provide sufficient differences in temporal envelopes across the various signals to facilitate segregation (cf. Exp. I). Possibly, across-frequency integration based on temporal envelope coherence helps to facilitate segregation (cf. Trahiotis and Stern 1994), but within-channel cues may also improve even when the signal exceeds one critical band.

One speculation would be that the monaural temporal envelope cues somehow help to select temporal intervals that belong to one of the two signals and that in this way temporally varying binaural cues are organized to facilitate segregation.

Alternatively, the auditory system may adopt 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 may help to tip the balance towards one of the hypotheses based on the assumption that within one auditory stream, binaural cues have to be coherent across time.

In a previous contribution (van de Par *et al.* 2005) we reported a study were listeners had to discriminate between BPN and HTC signals both presented interaurally out-of-phase within an in-phase noise masker. Listeners were able to discriminate between these signals at signal-to-masker levels for which no monaural cues were available for performing this task. This suggests that within the binaural display, information is available about the temporal structure of the out-of-phase signal. An equalization-cancellation (EC) stage (e.g. Durlach, 1963; Breebaart, van de Par, and Kohlrausch 2001) could in principle provide such information, because it would cancel the noise masker component of the stimulus but not the out-of-phase signal component. It would require, however, that within the binaural display, the capacity to process temporal information is sufficiently good to distinguish between HTC and BPN signals.

In the current study, in a similar way, an EC stage could remove one of the two signals, allowing for the temporal processing of the other signal. It is not evident, however, how this could explain the finding of Exp. III, that the in-phase AM BPN could not be segregated from the HTC. Possibly it is related to the observation that for any internal time delay that is applied in the equalization stage, the output of the cancellation stage has one dominant modulation rate of 20 Hz. For the out-of-phase condition, however, modulation rates of 40-Hz and 20-Hz can be observed depending on the internal delay. This difference in modulation rate across the binaural display may facilitate segregation.

Finally, we want to draw attention to the rather low lateralization threshold of 29 μs that was found in the ITD discrimination conditions with wideband stimuli. This indicates that a difference in azimuth between the simultaneously presented HTC and the BPN of about 6 degrees in a free-field condition would already be sufficient for listeners to notice the difference in azimuth.

In conclusion, this study presents an interesting stimulus paradigm that reveals binaural segregation based on monaural temporal envelope cues. Results are not easily understood in terms of existing binaural models.

References

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, pp. 1074-1088.

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, pp. 785-797.

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

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

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