Principles of Auditory Object Formation: On the influence of Spatial Information on Onset Detection

Othmar Schimmel[1], Armin Kohlrausch[1,2]

[1] *Technische Universiteit Eindhoven, The Netherlands, Email: o.v.schimmel@tm.tue.nl*

[2] *Philips Research Laboratories Eindhoven, The Netherlands, Email: armin.kohlrausch@philips.com*

Abstract

In a temporal positioning experiment, the influence of spatial information on onset detection of a short wide-band noise target in a wide-band noise masker was explored. Significant differences between the measured conditions were obtained, and an influence of spatial information was established.

Introduction

The ability to segregate and identify sound sources in an auditory scene comes naturally to most people. In *Auditory Scene Analysis*, the human listener's ability to group signal components, and consequentially separate discrete sources from a complex mixture of sounds, is the focus for research on perceptual stream segregation, or auditory object formation [1]. A new auditory object is initiated by detection of an increase of intensity, often occuring simultaneously over multiple spectral parts, the so-called onset. These spectral signal components are then grouped to form a new auditory object. The auditory system becomes more robust in grouping signal components within complex acoustic scenes by making use of interaural differences, the main cues for perceiving the spatial position of a sound source [2].

Since interaural disparities are inherent to every sound and they are already present at the onset, it is not a priori obvious whether they are the *cause* of grouping, or the *effect* [3]. From the available literature, it is difficult to decide whether localization cues support grouping of signal components to detect the onset of a new auditory object, or whether grouping of signal components by onset detection is a prerequisite for localization of the auditory object. In this research context, we try to answer the question: What is the influence of directional information on onset detection?

Experiment

Method

The experiment described here was defined to systematically determine the influence of lateralization cues on onset detection of wide-band noises, for their contribution to grouping of signals into separate auditory objects. The aim was to measure whether specific differences in interaural parameters between an existing sound and a new emerging sound influence the ability to perceive the onset of the new sound, and facilitate consequential stream segregation. This was achieved by varying the directional information of a target sound presented within a masking sound, and measuring the accuracy of temporally aligning the target sound to the meter of a series of isochronous marker pulses, in a procedure adapted from P-center temporal positioning experiments [4]. The time difference between the temporal position set by the subject from the time instant of a non-present third marker pulse in a series of five marker pulses serves as a quantitative measure for onset detection.

The hypothesis was that, for different spatial cues, by increasing the signal-to-noise ratio between target and masking sound, up from just below the target's detection threshold, detection of the target sound onset improves as a result of binaural processing. Alignment of this onset to the temporal center between the second and fourth marker pulse onsets then becomes more accurate, resulting in smaller time deviations. By choosing target levels relative to the threshold per spatial condition and per subject, the assumption is that the data can be pooled across subjects, and possibly an effect of spatial information on onset detection can be observed.

Three types of auditory objects were used: a continuous background masking sound with a level of 70 dB, a number of 50-ms marker pulses at 76 dB, and a 50-ms target sound at five levels around its detection threshold in the masking sound. All types were independent Gaussian noises, with immediate onsets and offsets. The masking sound, as well as the marker pulse, was identical at both ears, to establish lateralization in the center. The target sound was set at two positions in the lateral plane, one in the center and one to the far right, by manipulation of Interaural Time Differences, Interaural Level Differences, combinations of ITDs and ILDs, and filtering using Head-Related Transfer Functions.

Since there is no physical difference in the acoustical signal for lateralization in the center for both interaural time and level differences, the following six conditions were presented: ITD [center], ITD [right], ILD [right], a combination of ITD and ILD [right], HRTF [center] and HRTF [right]. The parameter value of the ITD [right] condition was estimated from [5] to be a 660 μs delay at the left ear. The value of the ILD [right] condition was also estimated from [5] to be 18 dB; lateralization was realized by increasing the target level 9 dB at the right ear, and decreasing it 9 dB at the left ear. The HRTF [center] and HRTF [right] conditions were computed from the head-related impulse responses at respectively 0 and 90 degrees of the HEAD acoustics HMS II artificial head [6].

Five different sound pressure levels were applied to the target sound: from -1 to +3 dB, relative to the detection threshold of the target sound, which was measured per condition per subject in a three-alternative forced-choice transformed up-down procedure [7].

The six spatial conditions, at five different target levels with masking sound and one target level without masking sound as a control condition, were randomly presented, over headphones, twice to ten subjects (N=20). The time deviation of the adjusted target position from the bisection of the second and fourth marker pulse onsets was measured. Since we were interested in a measure for the variability of this time adjustment, we used the absolute value of the time deviation for the further analysis.

Results

Figure 1 represents the results for all six spatial conditions. Plotted per condition are the means of the absolute time deviations from the temporal center, for the five target levels relative to the detection threshold for that particular condition with masking sound (-1 to +3 dB), and one target level without masking sound (Inf). The results demonstrate that an increase in signal-to-noise ratio between target and masking sound yields a decrease in the mean of the absolute time deviation. When no masking noise is present, the time deviations are the lowest. The overall time deviations in both [center] conditions are lower than in the [right] conditions.

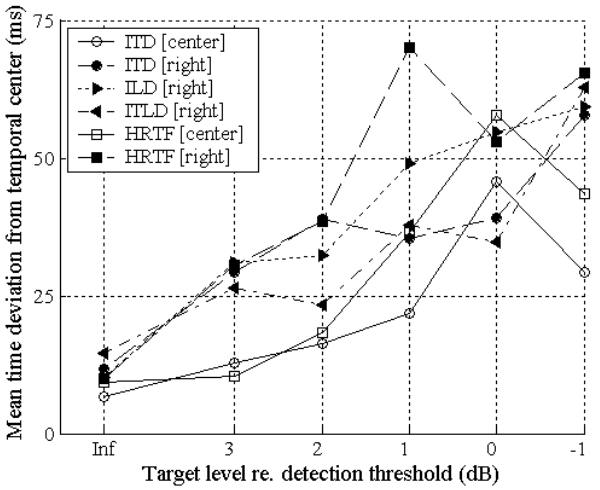

Figure 1: Mean of the absolute time deviation from the temporal center between marker pulse onsets versus target level relative to its detection threshold (N=20).

Discussion

Analysis of variance shows, besides an expected significant difference between target levels (p=0.000), that the difference between the six spatial conditions is significant (p=0.000). There is, however, no significant difference at the 5% level between the two [center] conditions (p=0.079), and between the four [right] conditions (p=0.140). Paired comparison of ITD [center] with each [right] condition supports this notion (p=0.027 or lower), as does paired comparison between all [right] conditions (p=0.052 or higher). Paired comparison of HRTF [center] reveals only significant differences with ILD [right] (p=0.026) and HRTF [right] (p=0.010), while for ITD [right] (p=0.175) and ITLD [right] (p=0.480) no significant difference is obtained. There is no significant interaction between spatial condition and target level (p=0.646).

By maintaining an equal internal signal-to-noise ratio, due to relative levels per condition, these findings might indicate an effect of binaural processing for the [right] conditions. For these conditions, with large disparities between both ears, accurate onset detection seems to be more difficult. Apparently, binaural processing of larger differences between left and right ear, combined with unmasking the target sound from the masking sound, interfers with determining the exact start of the target sound, even though the onset is immediate.

In this experiment, the signals in the different conditions were only compared for the same *sensation level*. Since time deviations in the [right] conditions, which require binaural unmasking, are higher than in the [center] conditions, it might be possible that the values would agree better when the signals are compared for the same *absolute level*.

Open to discussion are the choices for the ITD, ILD and HRTF for the [right] conditions. Since these involved both estimations ([5]) and measurements ([6]), it is likely that the directional cues did not match exactly between conditions. However, operationally they were not too different, since lateralization at the far right occured in all four conditions.

Conclusion

An influence of spatial information on onset detection is established: The significant differences between conditions with different interaural disparities suggest heavier binaural processing to be responsible for deteriorated performance in detecting the target sound onset.

References

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

[2] Blauert, J. (**1983**). *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, Cambridge (MA).

[3] Bregman, A.S. (**1998**). Psychological Data and Computational ASA. Rosenthal, D.F. and Okuno, H.G. (eds.) *Computational Auditory Scene Analysis*. Erlbaum, Mahwah (NJ), 1-12.

[4] Pompino-Marschall, B. (**1989**). On the psychoacoustic nature of the P-center phenomenon. *Journal of Phonetics* **17**, 175-192.

[5] Grantham, D.W. (**1995**). Spatial Hearing and Related Phenomena. Moore, B.C.J. (ed.) *Hearing. Handbook of Perception and Cognition* (2nd ed.), Academic Press, San Diego (CA), 297-345.

[6] Blauert, J., Brüggen, M., Hartung, K., Bronkhorst, A., Drullmann, R., Reynaud, G., Pellieux, L., Krebber, W., and Sotteck, R. (**1998**). The AUDIS Catalog of Human HRTFs. *Proceedings of the 16th International Congress on Acoustics*, 2901-2902.

[7] Levitt, H. (**1971**). Transformed Up-Down Methods in Psychoacoustics. *Journal of the Acoustical Society of America* **49**, 467-477.