SEGREGATION OF TWO SIMULTANEOUSLY ARRIVING NARROWBAND NOISE SIGNALS AS A FUNCTION OF SPATIAL AND FREQUENCY SEPARATION

Toni Hirvonen

Laboratory of Acoustics and Audio Signal Processing
P.O. Box 3000, FI-02015 HUT Finland
Toni.Hirvonen@hut.fi

ABSTRACT

The present paper details a set of subjective measurements that were carried out in order to investigate the perceptual fusion and segregation of two simultaneously presented ERB-bandlimited noise samples as a function of their frequency separation and difference in the direction of arrival. This research was motivated by the desire to gain insight to virtual source technology in multichannel listening and virtual acoustics applications. The segregation threshold was measured in three different spatial configurations, namely with a 0°, a 22.5°, or a 45° azimuth separation between the two noise signals. The tests were arranged so that the subjects adjusted the frequency gap between the two noise bands until they in their opinion were at the threshold of hearing two separate sounds. The results indicate that the frequency separation threshold is increased above approximately 1.5 kHz. The effect of angle separation between ERB-bands was less significant. It is therefore assumed that the results can be accounted by the loss of accuracy in the neural analysis of the complex stimulus waveform fine structure. The results are also relatively divergent between subjects. This is believed to indicate that sound fusion is an individual concept and partly utilizes higher-level processing.

1. INTRODUCTION

Two interaural cues, the interaural time and level differences (ITD and ILD, respectively), have been widely accepted to be mainly responsible for the perception of the left/right direction of a sound event [1]. If a pointwise sound source emits wideband sound in anechoic conditions, the produced localization cues indicate the same source direction at all frequencies. The present author has recently been studying the perception of sound events where the interaural cues indicate different source directions as a function of frequency [2][3]. Such conditions are not purely theoretical: sometimes applying different spatialization techniques in multichannel reproduction results in contradicting localization cues [2].

A simple way to simulate interaural cues that change unnaturally as a function of frequency is to route non-overlapping frequency bands to loudspeakers with different azimuth angles, and play the signals from different speakers simultaneously. It is not well known how these sound events are perceived. Previous studies with noise signals indicated that in cases where several adjacent frequency bands with different azimuth directions were presented, the subjects never perceived the sound from all loudspeakers [3]. Thus, some noise bands, even though not overlapping in frequency or spatially, were perceptually fused together.

A possible explanation for the previous result is that the components of a complex signal are associated as being fused to one sound when they are closely related in terms of some perceptual attribute. These attributes may include frequency content, arrival time (precedence effect), temporal structure, spatial direction etc. When the difference between components in one or more of these attributes increases, it is more likely that the contrast causes separate sounds to be heard. This theorem follows the general lines of Gestalt theory, which is commonly utilized in perceptual psychology. However, it requires at least some amount of processing at the higher stages of perception. Similar hypotheses have also been presented in other studies [4].

The goal of this study is to help understand the perceived fusion and segregation of noise samples in anechoic conditions. More precisely, the measurements were concerned with the necessary frequency gap between two one-ERB noise band components that causes two separate sounds to be heard, as a function of frequency and spatial angle. The fusion and segregation of virtual sound sources are important phenomena in multi-channel audio reproduction and virtual acoustics applications. The results could possibly help understand the perception of sound source width as well. It should be emphasized that the precedence effect is not considered in this study, as the noise components were arranged to arrive simultaneously.

2. PROCEDURE

2.1. Test Method

Segregation was measured by listening to two simultaneously arriving noise samples. The bandwidth of both samples was one ERB, calculated according to the equivalent rectangular bandwidth scale [5]. The segregation threshold value was determined by adjusting the frequency of the higher ERB-band noise component, while the lower-frequency noise remained constant.

The samples in the test were created by first filtering white 80 dB Gaussian noise according to the equal loudness contour measured by Robinson and Dadson [6]. The equal-loudness noise was then divided into bandpassed, one-ERB-wide samples. The filtering was done via FFT, so that the filters' magnitude response passbands were very close to rectangular. The length of each sample was one second with a 20 ms exponential onset and offset.

During the listening tests, two ERB-noise samples were listened to simultaneously in each case. The case was looped for the duration of the adjustment, so that the one second of sound was followed by a 400 ms silence. At the beginning of each test case, the two samples were presented in adjacent ERB-bands, i.e. the frequency gap between the samples was zero. The subjects were instructed to increase the frequency gap to the point where they

clearly heard two separate sounds, and to go back and forth for a few times between the sensations of one and two sounds. After a few adjustments, the subjects determined the proper threshold value somewhere inside this range and proceeded to the next case.

The subjects changed the frequency gap to both directions in two step sizes. In this paper, all adjusted frequency gap values that are presented in ERB units refer to the ERB bandwidth that is calculated according to the center frequency of the lower noise component. The large step increased or decreased the gap by the amount of one ERB. The small step size was similarly one-fourth of an ERB. However, the adjustment had a distinct lower limit: the test system did not allow for the two bands to overlap in frequency to avoid interference. Also, an upper limit of 10 ERBs was used in case the two components would not segregate at all. It should be emphasized that the upper frequency component's bandwidth was re-calculated and changed after each adjustment step, so that the subject always listened to two ERB-wide samples.

This procedure was repeated in 13 different values of the lower noise component's center frequency. These values spanned the range of 0.1-5kHz in 2 ERB increments, according to the ERB scale. Furthermore, these 13 cases were repeated with three different azimuth angles between the two frequency components: 0°, 22.5°, and 45°. The lower component always emitted from a loudspeaker directly in front of the subject (0°) Thus, the total number of test cases was 39.

The subjects performed the listening test in an anechoic chamber. The subjects were seated in a chair facing 0° and told not to move their heads during tests. The test utilized three loudspeakers at 0°, 22.5°, and 45°. The loudspeakers were located approximately 2 m from the listener's head at eye-height horizontal plane. The loudspeaker distances from the listening spot were compensated with delays, so that the two components arrived at the same time. The loudspeakers' magnitude responses had been measured and equalized to be flat within 1 dB in the listening position. Loudspeaker levels were also aligned using wideband noise and a sound level meter. Figure 1 illustrates the test setup in the anechoic chamber.

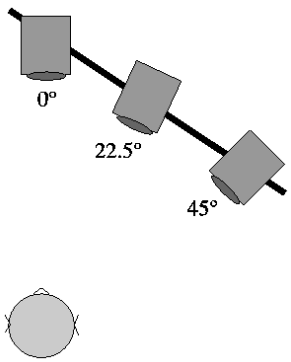

Figure 1: *The loudspeaker setup in the anechoic chamber used in the listening test. Subjects listened to two simultaneously arriving, non-overlapping ERB-band noise components, whose azimuth separation was either 0°, 22.5°, or 45°. In each test case, the subjects adjusted the frequency of the higher component until they heard two separate sounds.*

2.2. Test Subjects

Prior to the formal experiments, it was necessary to ensure that the subjects both understood what they were required to do, and were all evaluating the same phenomenon. For this reason, a prescreening experiment was arranged. In the experiment, the subject candidates listened to diotic samples with headphones that contained two ERB-band noise components, similarly as in the actual listening test. Among the cases were two that contained a similar lower frequency component, centered at 250 Hz. The difference between these cases was that the gap between the lower and higher component was 0 (i.e. one continuous two-ERB sample) in the other and 10 ERB in the other.

Preliminary screening required that the subject candidates reported whether they heard one or two separated sounds in diotic listening. The majority of subjects that attended reported that they heard only one sound when the gap was zero and two distinct sounds when the gap was 10 ERB. Two candidates out of eight were excluded from the tests based on their prescreening performance: one reported hearing both above cases as one sound, whereas the other reported both containing "many sounds". Thus, six persons participated in the final test.

The exact question in the test was to find the threshold of hearing two separate sounds. The task was further discussed so, that the subjects were asked to focus on hearing a distinct interval and to keep in mind that "separate sounds" and "two sound sources" are not the same thing: one can perceive two distinct sounds coming from the same loudspeaker.

It must be emphasized that the listening tests were conducted to relatively compare fusion/segregation as a function of frequency and azimuth angle. The prescreening was done in order to find subjects that experienced the fusion/segregation phenomenon similarly in a natural manner, rather than trying to train or bias subjects into hearing something that they did not naturally hear. The role of individual differences in the listening task is discussed after the results are presented.

3. RESULTS

Panels 1-6 of Figure 1 illustrate the results from the listening tests individually for each subject. Each panel shows three data curves, which represent the three different spatial angles between the two noise components, whose segregation threshold was measured. As mentioned, the subjects adjusted the center frequency of the higher frequency component as the lower component remained constant in each case. The x-axis frequency value represents the center frequency of the constant lower frequency band, whereas the y-axis gives the frequency gap threshold between the two components required for the subject to hear two separate sounds. The gap threshold value is given in ERBs, whose bandwidths are calculated according to the center frequency of the lower component in each case. This way, the segregation threshold values can be compared at different frequencies and spatial angles.

The results have somewhat large inter-subject variance. However, one salient feature is common to all individual results: the segregation threshold is notably increased in cases, where the center frequency of the lower component is above approximately 1.5 kHz. The significance of this phenomenon varies from subject to subject. Subjects 1,2, and 5 have marked the threshold at high frequencies to be at or near the upper adjustment limit of the test system. The influence of the spatial angle between the two fre-

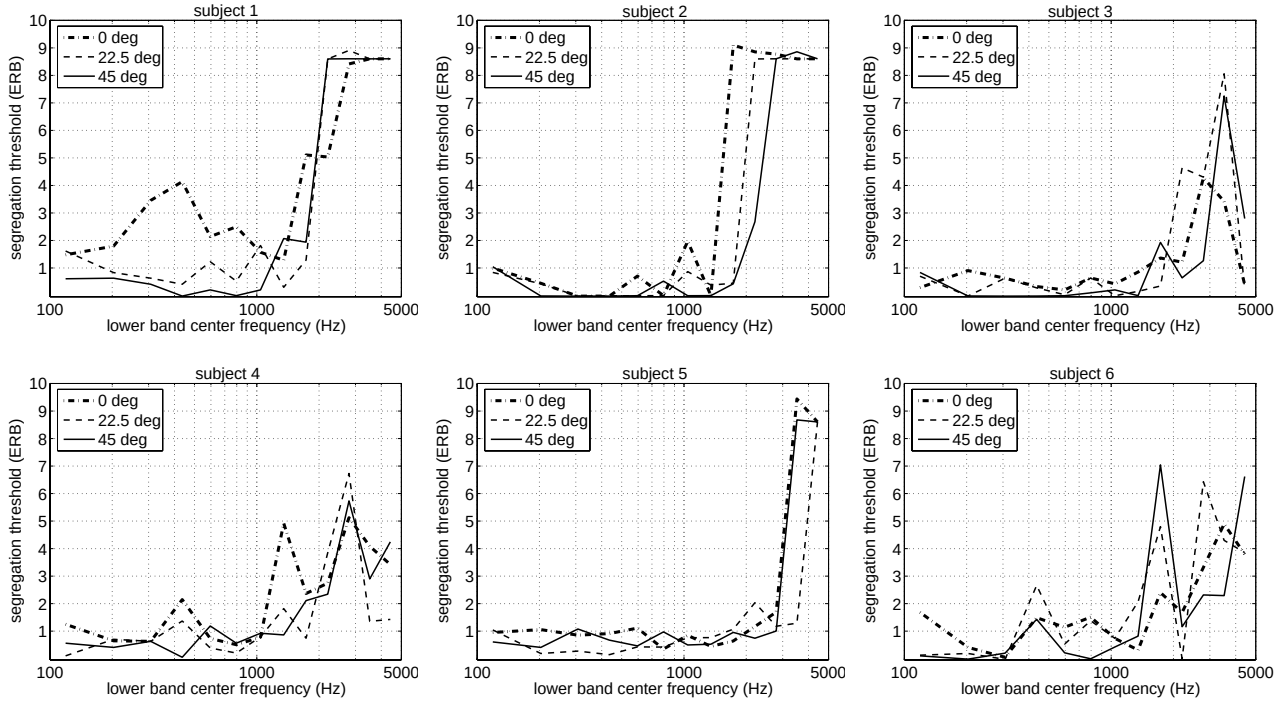

Figure 2: *The results of the listening tests presented individually for each subject. The three curves represent the segregation threshold values with three different azimuth separations. The x-axis frequency value represents the center frequency of the constant lower ERB-band component, whereas the y-axis gives the frequency gap threshold between the two components required for the subject to hear two separate sounds. The gap threshold value is given in ERBs, whose bandwidths are calculated according to the center frequency of the lower frequency band in each case.*

quency components is less prominent and hard to interpret from the results.

4. DISCUSSION

In this section, the previous results are analyzed further and conclusions drawn from them. Figure 3 shows the mean thresholds for the three component separation azimuth values averaged across all six subjects. The mean threshold is approximately 1 ERB below 1.5 kHz. It can be seen that the threshold value begins to increase when moving above 1.5 kHz in all three cases. Although the mean threshold is a little higher for 0°, the effect of the azimuth separation between the two ERB-bands is difficult to determine from the results. Sufficient to say however, that the azimuth angle difference did not contribute to the segregation threshold value as strongly as the frequency content of the stimuli. It is therefore reasonable to assume that the segregation task mainly utilized monaural processing, and is not as much dependent on binaural factors, such as interaural localization cues. There are many well-known cases, where the accuracy of hearing is decreased with increasing frequency, such as the perception of pitch [7] and waveform ITD [1] [8]. The lack of sensitivity at high frequencies is often attributed to the loss of phase locking in the firing patterns of the cochlear auditory nerve fibers [9]. The general view is that the neural synchrony with the sound waveform features begins to decline already at relatively low frequencies and is completely lost above 5 kHz [7]. This in turn implies that the exact time structure of a complex waveform is not efficiently encoded at high frequencies.

In the light of the previous facts, a plausible hypothesis is that the segregation threshold for the two noise bands is lower when the complex stimulus waveform can be accurately encoded by the neural system. Figure 4 shows an example stimulus, similar to those used in the listening test, which consists of two ERB-band components (center frequencies 120 and 690 Hz). Because this stimulus waveform can be analyzed accurately, the two components are more likely to be perceptually segregated. In order to illustrate the perceptual segregation, the lower panel shows the waveforms of the two components separately. When both noise band components of the stimulus are instead above 1.5 kHz, only the overall lower-frequency envelope of the stimulus can be detected accurately, and the segregation threshold of the two components increases. In some higher-frequency cases the subjects were not able to segregate the two bands even after reaching the upper 10-ERB limit of the test system.

In the course of the tests it became clear that the fusion/ segregation question is also a subjective one. Based on prescreening verbal comments, humans might consider various factors, such as harmonic relations, presence of an interval, and spatial difference, when determining the perceptual components of a sound. In this experiment, the subjects were instructed to use the perceived presence of an interval as an indication of segregation. The prescreening indicated that this could be one of the principal factors according to which the majority of humans associate segregation naturally. The purpose was, however, to compare the segregation thresholds relatively as a function of frequency and spatial angle,

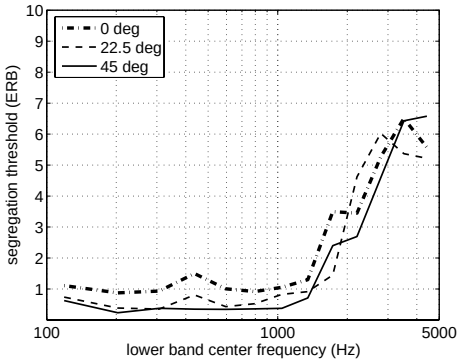

Figure 3: *The mean segregation threshold values averaged across all six subjects, presented similarly as in Figure 2.*

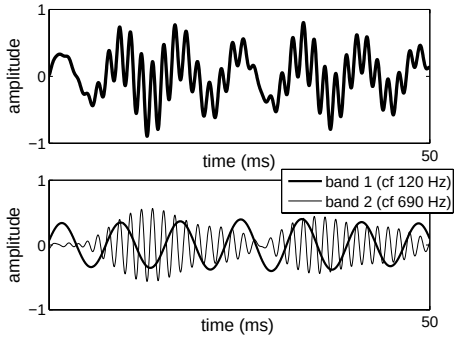

Figure 4: *Upper panel: example stimulus in time-domain. Two ERB-bandwidth noise components with center frequencies (cf) 120 and 690 Hz are presented simultaneously. Lower panel: below 1.5 kHz, the signal fine structure is accurately analyzed and the individual waveforms of the two noise bands can be segregated into separate sounds. Above 1.5 kHz only the common envelope is perceived.*

instead of defining the different perceptual concepts. The prescreening test was arranged so that subjects would, at least roughly, "measure" the same phenomenon.

Nevertheless, as can be seen from the results, there are deviations between the subjects. For example, subjects 3, 4, and 6 did not reach the upper adjustment limit at high frequencies, whereas the other three did. It is also evident that these same subjects' results are not as consistent with frequency as the others' are. It is possible that despite the prescreening, the subjects did not perform the task using exactly the same criteria. It seems that the fusion/segregation task included at least some amount of higher-level processing that extends beyond peripheral hearing. For these reasons, it was chosen to present the results individually, and not to increase the number of test subjects.

The individual deviations withstanding, the change in behavior when moving to the higher frequency range can be clearly observed from the results. One specific task left to the future is to establish whether the present results can be utilized in auditory modeling of perceived source width. Based on this study, the initial hypothesis is that in wideband sound, where interaural cues are not consistent, frequency components in the range above 1.5 kHz in steady-state sound contribute less to perceived source width.

5. CONCLUSIONS

The perceptual segregation and fusion of two simultaneously arriving ERB-bandwidth noise components was investigated as a function of their frequency- and azimuth separation in an anechoic chamber. The subjects adjusted the frequency gap between the components until they heard two separate sounds. The results indicate that when both components are above 1.5 kHz in frequency, i.e. the range where fine structure information of a complex stimulus is more difficult to determine, the frequency gap threshold is notably increased. The present hypothesis is that at high frequencies, where only the signal envelope can be analyzed accurately, the components are more difficult to segregate. The effect of azimuth separation between the ERB-bands was not prominent compared to the frequency dependency. Although the main trends can be observed easily from the results, the inter-subject deviations were relatively large. It is hypothesized that determining the segregation of two frequency components utilizes at least some amount of higher-level processing.

6. ACKNOWLEDGMENTS

The author wishes to thank Prof. Matti Karjalainen and Dr. Ville Pulkki for their suggestions. This work has been supported by The Academy of Finland (project no. 201050).

7. REFERENCES

[1] Blauert, J., Spatial Hearing, Revised edition MIT Press, Cambridge, MA, USA, 1997.

[2] Pulkki, V. and Hirvonen, T., "Localization of Virtual Sources in Multichannel Audio Reproduction", IEEE Trans. SAP, Vol. 13, no. 1, pp. 105-119, 2005.

[3] Hirvonen, T and Pulkki, V., "Perception of Frequency-Dependent Interaural Cues", unpublished.

[4] Gardner, M.B., "Image Fusion, Broadening, and Displacement in Sound Localization", J. Acoust. Soc. Am. Vol. 46, No.2, pp. 339-349, 1969.

[5] Glasberg, B.R. and Moore, B.C.J., "Derivation of Auditory Filter Shapes from Notched-Noise Data", Hear. Res., vol. 47, pp. 103-138, 1990.

[6] Robinson, D.W. and Dadson, R.S., "A Re-Determination of the Eq Loudness Relations for Pure Tones," Brit. J. Appl. Phys., vol. 7, pp. 1 181 (1956).

[7] Hartmann, W.M., Signals, Sound, and Sensation, AIP Press, 1997.

[8] Henning, C.B., "Detectability of Interaural Delay in High-Frequency Complex Waveforms", J. Acoust. Soc. Am. Vol., 55, No.1, pp. 84-90, 1974.

[9] Palmer, A. R. and Russel, I.J., "Phase-Locking in the Cochlear Nerve of the Guinea Pig and its Relation to the Receptor Potential of the Inner Hair Cells" Hear. Res., 24, pp. 1-15, 1986.