7TWO-POINT DISCRIMINATION IN AUDITORY DISPLAYS

Virginia Best[1] *, André van Schaik*[1,2] *and Simon Carlile*[1]

1 Department of Physiology (F13) and 2 School of Electrical and Information Engineering
University of Sydney
Sydney, NSW 2006, Australia
{ginbest,simonc}@physiol.usyd.edu.au and andre@ee.usyd.edu.au

ABSTRACT

In this paper we describe work which characterises the effect of spatial factors on the segregation of concurrent sound sources. The results inform the operational requirements of virtual auditory displays required to render multiple, concurrent sound sources in terms of (i) minimum spacing between sources and (ii) identification of the principal acoustic directional cues exploited by the auditory system for source segregation. Experiments using various broadband sound sources (white noise, click trains, spoken words) indicate that the extent of actual separation required for reliable segregation of concurrent stimuli varies as a function of location. The pattern of location dependence indicates that the auditory system is principally exploiting binaural differences for sound segregation. Monaural spectral cues, while essential for high fidelity spatialisation, seem to play a much less minor role in segregation under these conditions. However, spectral cues are likely to be useful when competing stimuli have distinct temporal structures or are not fully coincident in time.

1. Introduction

In the process of rendering multiple concurrent sound sources in a virtual display, the issue of separability of information streams is very likely to be critical in the successful application of the display. In normal listening in the free field, humans are capable of clearly separating multiple sound sources and streaming or parsing the information (say speech) from each source location, sometimes under remarkably complex acoustic conditions. A clear understanding of the limits of this process, together with the way in which source characteristics interact (e.g. spectral content and spatial locations) and the acoustic cues exploited by the auditory systems in managing this process all inform the minimum operational requirements for auditory displays.

Differences in the spectral content of temporally overlapping sounds provide strong cues to the auditory system for streaming out the information from different sources [1,2]. For instance, different voices generally differ in pitch and can be parsed on this basis at some early stage of auditory scene analysis. Disparities in the temporal overlap can also aid in the separation of competing sound sources, where even a brief glimpse of a target in isolation from distracting sources significantly reduces the masking effect of the competitor [3,4]. It has also been known for some time that the relative location of concurrent sound sources has an impact on their separability however, less is known about the processes mediating this advantage [see 5 for review].

In the auditory system, the spatial origin of a sound source must be computed from a number of acoustic cues (this is in contrast to other sensory systems, such as vision, where a two-dimensional representation of space is directly encoded by the sensory epithelium). The binaural cues arise due to the placement of the two ears on either side of the head. For a sound source located away from the midline, the difference in path length to the two ears results in a difference in the arrival time of the sounds at each ear (interaural time difference; ITD). Furthermore, a sound originating from one side of the head will be more intense in the near ear than in the far ear, resulting in an interaural level difference (ILD). As the ears are placed more or less symmetrically on the head, the binaural cues are ambiguous and describe a set of locations on a cone centred on the interaural axis- the so called "cones-of-confusion". To resolve these ambiguities, the auditory system relies on small head movements as well as the three-dimensional shape of the external ears and head, whose location-dependent spectral filtering provides monaural cues to sound source location. These cues, in combination, allow a robust estimation of the location of a single sound source.

It is not clear, however, how robust this computational coding of auditory space is in the context of multiple concurrent sound sources. The addition of extra sources to the auditory scene causes direct interference between the sources because they all share the same computational channels. The aim of this work has been to examine the perceptual resolution of the auditory system for pairs of concurrent auditory stimuli using a two-point discrimination paradigm. This work has also examined the principal acoustic cues used by the auditory system in this discrimination.

2. Methods

2.1. Virtual Auditory Space Stimulation

Individualised Virtual Auditory Space (VAS) is generated by accurately simulating the wave pattern at the eardrum occurring after free field stimulation with an external sound source [5]. The procedure employed in our laboratory for generating VAS has been described elsewhere in detail [6,7]. Briefly, directional transfer functions (DTFs) are obtained for a large number of locations in space by recording impulse responses from miniature microphones (Sennheiser KE 4-211-2) placed in the ear canals of an individual. The DTF filters for each ear corresponding to a particular spatial location can then be convolved with any stimulus and delivered via in-ear tubephones (Etymotic Research ER-2) to the subject to give rise to a virtual externalised sound stimulus. To generate stimuli imitating a concurrent sound source pair, each stimulus was filtered with the DTFs corresponding to its desired location before addition of the pair. For "zero separation" stimuli, the same procedure was followed but the same DTFs were used to filter both stimuli of the pair.

2.2. Sound Stimuli

As the aim of this work was to examine the spatial representation of multiple sounds, it was important to choose stimuli that are well-localised when presented independently. Two uncorrelated broadband white noise bursts (300 Hz to 16 kHz) were played, and the subject was required to indicate if the sources were located at the same or different locations. Under these conditions, the two sounds to be separated were indistinguishable on the basis of pitch, timbre or content, allowing an independent investigation of the effects of location on segregation.

2.3. Testing Procedure

Four normal hearing listeners participated in the experiment, and were seated in a darkened, sound-attenuating chamber for testing. Sounds were presented at approximately 50dB sensation level and subjects indicated their perception by pressing one of two buttons on a response box (e.g. left button - both sources at the same location; right button - sources from different locations). The locations of the test stimuli are described in terms of a single pole coordinate system. Testing took place using five reference locations on a horizontal plane at the level of the audio-visual horizon (0 ° elevation): 0 °, 22.5 °, 45 °, 67.5 ° and 90 ° azimuth on the right hand side of the subject (Figure 1). In each trial, one stimulus in the concurrent pair was presented from one of the reference locations and the other from one of 15 test locations displaced in either azimuth (horizontal separation) or elevation (vertical separation). For vertical separation, the testing range spanned from -45 ° to 90 ° elevation in 10 ° steps. Ranges for horizontal separation varied with location to cover a suitable range: ±21 ° at 0 ° azimuth; ±32 ° at 22.5 ° azimuth; ±42 ° at 45 ° azimuth; ±53 ° at 67.5 ° azimuth; ±63 ° at 90 ° azimuth. Note that the binaural cues change with horizontal separation and also with vertical separation (off the midline) using this coordinate system. All reference location and separation combinations were presented 10 times each in a random order.

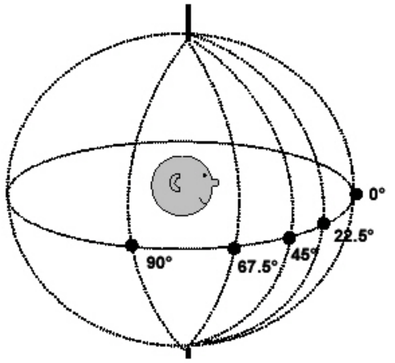

Figure 1: *The single pole co-ordinate sphere used to define locations in space. The black dots illustrate the five reference locations examined - 0, 22.5, 45, 67.5 and 90° azimuth, all on the 0° elevation plane at the level of the ears.*

3. RESULTS

The psychophysical curves were plotted for each subject and each reference location. Figure 2 shows the data for one subject for both vertical and horizontal separation of the targets.

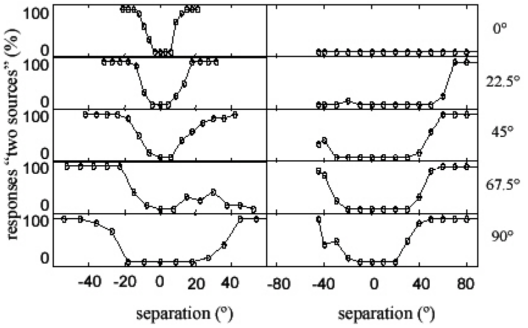

Figure 2. *Psychophysical data for one subject. The left column shows data for horizontal separation and the right shows vertical separation. The Y- axis of each panel extends from 0 to 100% and the percentage of subject responses indicating the perception of two sources is plotted against actual separation.*

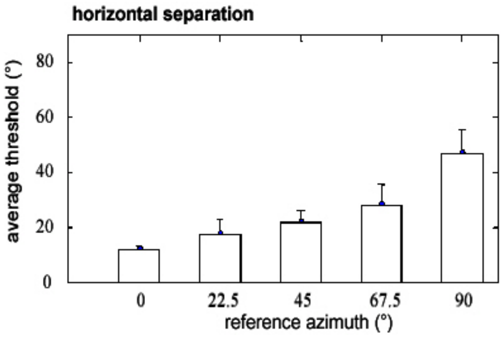

Figure 3. *Average horizontal separation thresholds (with standard deviations) across subjects for the 5 reference locations.*

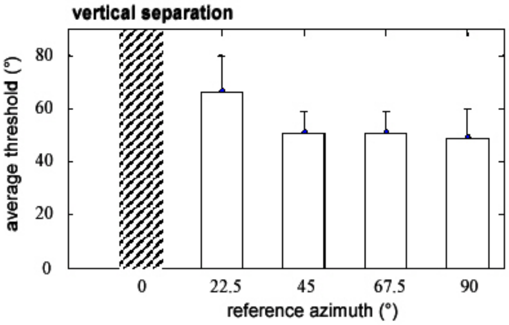

Figure 4. *Average vertical separation thresholds (with standard deviations) across subjects for the 5 reference locations. At 0 °, the hashed region indicates that no thresholds could be obtained.*

Thresholds pooled for all subjects were calculated from the psychophysical curves and defined as the separation value at which both sources were perceived in 75% of trials. Figures 3 and 4 summarise the threshold data pooled for all subjects and all locations. Figure 3 shows the average angular thresholds and the standard deviation for stimuli separated horizontally about the reference location. At the most frontal location (0°azimuth), the two stimuli could be distinguished when separated by about 10°. Thresholds were seen to increase with more lateral positions of the stimuli and at the most lateral location (90° azimuth), about 45° separation was required to reliably distinguish two positions from one.

Figure 4 shows data for vertical separation plotted in the same manner as Figure 3. In Figure 4 the opposite trend in the data can be seen in that thresholds are high for more medial locations and decrease slightly at the more lateral locations. The most striking feature of these data was that at the frontal midline location, performance did not reach the 75% criterion over the entire testing range (-45 ° to 90 ° elevation) and hence no threshold could be obtained. This is indicated by hashed region in Figure 4.

4. ONGOING WORK

Ongoing experiments are taking a similar approach to the experiment described but using broadband speech stimuli. Concurrent pairs are made up of two different monosyllabic words of equal duration, spoken by the same male voice. These have also been band-passed (300 Hz to 16 kHz) and we have shown that they contain sufficient high-frequency information for accurate localisation [8]. These experiments differ from the main experiment as the stimuli are distinguishable on the basis of content, and accordingly a different task is employed in which subjects are required to assess the relative location of the two stimuli. For horizontal separation, they indicate whether a target word is to the right or left of the other word, and for vertical separation they indicate whether a target word is above of below the other word. Some preliminary data is shown in Figure 5, illustrating discrimination performance of one subject (the same subject as in Figure 2) at the front-most reference location (0,0). It can be seen that for horizontal separation (upper panel) responses were correct and consistent with just a few degrees of separation. However when the words were separated on the median vertical plane (lower panel) this ability was lost and the subject appeared to be guessing randomly even for the largest separation values. These results, although preliminary, indicate that the findings for concurrent broadband noise may well generalise to more complex broadband stimuli.

5. DISCUSSION

The data for horizontal separation on the 0 ° elevation plane are consistent with a predominant role for binaural cues in separation in this particular task. Due to the position of the two ears on this plane, binaural cues change as a sine function of azimuth, with the rate-of-change being maximum at 0 ° azimuth and decreasing towards zero at 90 ° azimuth. It follows that if separation is dependent on the difference in binaural cues between a pair of stimuli, then the threshold angle to achieve adequate binaural separation would increase with laterality. This is indeed the pattern seen in the individual data (Figure 2 left column) and in the mean data pooled across all subjects (Figure 3).

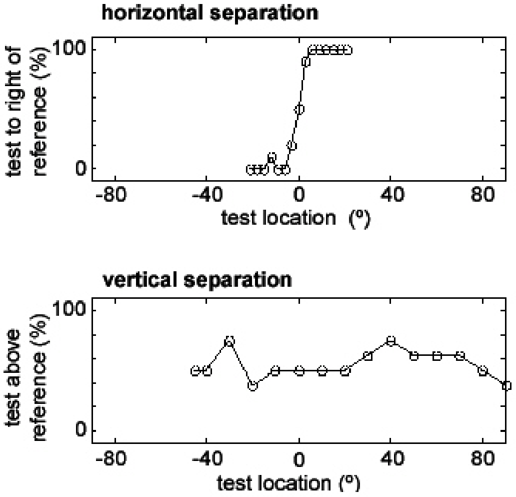

Figure 5. *Preliminary psychophysical data for one subject using spoken word stimuli. Upper panel - horizontal separation; Lower panel - vertical separation. The curves illustrate responses in a left-right (or up-down) discrimination task. The Y- axis of each panel represents the subject's certainty of the relative location of the two stimuli for the different separation values.*

These data are consistent with research on segregation using other kinds of stimuli, where binaural cues have been shown to offer an improvement. Previous work has demonstrated that simple stimuli such as tones are easier to detect in noise when the two are separated horizontally [9]. Similarly, the separation of speech and maskers along the horizontal plane decreases the masking effect of noise or other distracting talkers and renders the target speech more intelligible [10,11]. Furthermore, a recent study showed that the localisation of a broadband sound source in the presence of temporally interleaved distracters is less disrupted if the horizontal separation of target and distracters is large as compared to when they are placed at similar azimuths [12].

In contrast to the location-dependent pattern of change seen with horizontal separation, the thresholds for the vertical separation of concurrent sources were found to decrease with increasing laterality. This result is also consistent with the argument that the rate-of-change of the binaural cues is likely to be principally responsible for the pattern of responses seen. At the most lateral reference location (90 ° azimuth), vertical separation of a concurrent stimulus will cause a greater change in binaural cues than at more frontal locations. For instance, as the sound source is moved vertically from the interaural axis (90 ° azimuth) the displacement is towards the far ear and away from the near ear thereby decreasing the interaural disparity. However the strongest evidence that binaural cues are responsible for the performance of subjects is the fact that when stimuli were separated along the vertical midline, individual locations were not perceived by subjects (Figure 2, right column, top; Figure 4; Figure 5, lower panel). In this configuration, the binaural cues are effectively zero for both stimuli.

An important implication of the results obtained for the midline vertical separation is that spectral cues are not sufficient to indicate the presence of the distinct sources under conditions where there are multiple concurrent sources. This may at first seem surprising, since spectral cues are responsible for an extremely accurate ability to localise single noise sources on the vertical midline. For instance average localisation errors of approximately 5° at locations directly in front of the listener are reported for broadband noise [13,14]. However, it is possible that when two broadband sounds are presented concurrently, as in the current experiment, these spectral cues become less useful because the spectra of the sounds cannot be separated as they share the same frequency processing channels. Presumably, without the binaural cues to, in some way, group the spectral information associated with each source, this information simply interferes and results in a complete loss in the ability to perceive the presence of multiple sources on the midline plane. Most interestingly, in a separate set of experiments examining running speech, we have found that vertical separation of talkers in the median plane does increase the intelligibility of the target sentence [15]. It is possible that the auditory system makes use of unique temporal fluctuations and information content in running speech to segregate different sources. Interestingly, however, this effect was only seen when competing sentences were spoken by different voices, suggesting that prosody may be helpful in grouping spectra although they are utilising overlapping (if not identical) frequency channels.

6. SUMMARY

The novel finding reported in this study was that the ability of subject to separate concurrent multiple broadband sources seems to rely almost entirely on the presence of a binaural difference between the two sources. In spatial arrangements where only spectral cues are available (such as along the vertical midline), the individual locations cannot be perceived despite the fact that under conditions of single sound sources these cues provide the basis for relatively accurate localisation.

There are three principal implications of this work for the specification of auditory displays. Firstly, the relative positioning of auditory sources within the displays need to maintain a significant interaural cue disparity between sources if positional identity is an important information parameter. Secondly, the data presented here indicate that this separation is location-dependent. Thirdly, the correct rendering of binaural cues is critical to maintaining segregation performance where multiple concurrent sound sources are to be rendered.

7. ACKNOWLEDGEMENT

This work has been supported by the Australian Research Council and the National Health and Medical Research Council (Australia).

8. REFERENCE

[1] A. S. Bregman and J. Campbell, "Primary auditory stream segregation and perception of order in rapid sequences of tones," *J. Exp. Psychol.*, vol. 89, pp. 244-249, 1971.

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

[3] D. R. Perrott, "Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity," *J. Acoust. Soc.Am.*, vol. 75, pp. 1201-1206, 1984.

[4] M. A. Stellmack, "The reduction of binaural interference by the temporal nonoverlap of components," *J. Acoust. Soc. Am.*, vol. 96, pp. 1465-1470, 1994.

[5] S. Carlile, *Virtual Auditory Space: Generation and Applications*. Landes, Austin, 1996.

[6] D. Pralong and S. Carlile, "Measuring the human head-related transfer functions: A novel method for the construction and calibration of a miniature "in-ear" recording system," *J. Acoust. Soc. Am.*, vol. 95, pp. 3435- 3444, 1994.

[7] V. Best and S. Carlile, "Discrimination of sound source velocity in human listeners," *J. Acoust. Soc. Am.*, vol. 111, pp. 1026-1035, 2002.

[8] C. Jin, V. Best, S. Carlile, T. Baer and B. Moore, "Speech localization." In *Proc. Audio Eng. Soc. 112 th Convention*, Munich, Germany, May 2002.

[9] N. I. Durlach and H. S. Colburn, "Binaural Phenomena," in *The Handbook of Perception*, Eds. E. C. Carterette and M. P. Friedman. Academic, New York, 1978.

[10] I. J. Hirsh, "The relation between localization and intelligibility," *J. Acoust. Soc. Am.*, vol. 22, pp. 196-200, 1950.

[11] D. D. Dirks and R. H. Wilson, "The effect of spatially separated sound sources on speech intelligibility," *J. Speech Hear. Res.*, vol. 12, pp. 5-38, 1969.

[12] E. H. Langendijk, D. J. Kistler, and F. L. Wightman, "Sound localization in the presence of one or two distracters," *J. Acoust. Soc. Am.*, vol. 109, pp. 2123-2134, 2001.

[13] S. Carlile, P. Leong and S. Hyams, "The nature and distribution of errors in sound localization by human listeners," *Hearing Res.*, vol. 114, pp. 179-196, 1997.

[14] R. A. Butler, R. A. Humanski and A. D. Musicant, "Binaural and monaural localization of sound in two-dimensional space," *Perception*, vol. 19, pp. 241-256, 1990.

[15] V. Best, C. Jin, A. van Schaik and S. Carlile, "Spatial effects on speech intelligibility revisited," in *Proc. Australian Neurosci. Soc.*, Adelaide, Australia, January 2003, pp. POS-WED-269.