PARAMETERS FOR AUDITORY DISPLAY OF HEIGHT AND SIZE

Densil Cabrera and Steven Tilley

School of Architecture, Design Science and Planning
The University of Sydney
NSW 2006, Australia
densil@arch.usyd.edu.au

ABSTRACT

The vertical localization, and vertical and horizontal spread, of auditory images was investigated using a vertical 5-loudspeaker array, with the noise-band signal spectrum, loudness level, number of active loudspeakers, and center loudspeaker position as parameters. The signal spectrum and center loudspeaker position affected the image vertical center. Loudness level, and secondarily the signal spectrum, affected the auditory image size. These results, which are consistent with, and extend, previous studies into auditory volume and associations between stimulus frequency and vertical localization, suggest a robust basis for the auditory display of height and size.

1. INTRODUCTION

For an auditory display of data to function well, an intuitive relationship between the auditory and data parameters is desirable. Such a relationship should reduce learning time and increase the robustness of the display. This paper examines the possibility of auditory bases for auditory representation of *height* and *size*.

1.1. The Pitch-height Effect

The term 'pitch-height effect' is used here to refer to the tendency for stimuli with predominantly low frequency energy to be localized to positions lower than predominantly high frequency stimuli from the same source.

Whether this effect is merely a case of extrinsic association has been debated since Stumpf [1]. Associative bases are found in many languages (height describing pitch), singing pedagogy ('chest' and 'head' voice), loudspeakers (tweeters above woofers), acoustical terminology ('high' or 'low' frequency), sound sources (high frequency sources can be small and easily elevated, whereas ground vibration is associated with bass), and sound propagation (high frequency sound is easily blocked by objects on the ground, whereas low frequency sound diffracts well). Although such links combined may suggest a near-inevitable metaphor, exceptions are easily identified.

In an argument against inevitability, Zbikowski [2] notes that *sharpness* versus *heaviness*, *small* versus *large* and *young* versus *old* refer to what we know as 'high' versus 'low' musical pitch in ancient Greek, Balinese and the South American Suyá cultures respectively. He takes these exceptions as indicating a profoundly metaphorical basis for musical understanding of pitch. Scruton [3] argues in some detail that metaphors, exemplified by pitch-height, distinguish music from mere sound.

Initial experimental work on the pitch-height effect was done in the early 1930s. Using a small loudspeaker, Pratt [4] found that subjects localized five tones at octave intervals between 256 Hz and 4096 Hz to positions from low to high in frequency order. Although Dimmick and Gaylord [5] failed to duplicate Pratt's results, Trimble [6] found the pitch-height effect clearly evident in four out of five subjects for nine tones from 500 Hz to 3950 Hz. In the 1960s, Roffler and Butler [7] revisited the issue with more developed experiments, and confirmed the effect found by Pratt for nine tone frequencies from 250 Hz to 7200 Hz. They found the effect preserved for listener orientation (rather than the direction of gravity) when subjects lay on their backs or sides, and it was also found in congenitally blind subjects, as well as 4- and 5-year old children who appeared not to associate height and pitch linguistically. They also showed that the visual vertical scale used in experiments influenced the effect.

Using various band-limited noise and tone stimuli, Roffler and Butler [8] also examined the cues for vertical localization – concluding that accurate localization requires complex stimuli with energy above 7 kHz. Many other studies since have examined the role of spectral cues in median plane localization, identifying cues above 3 kHz due to pinna effects. With torso reflections included, cues can extend down to 700 Hz [9].

Localization of high frequency narrow band stimuli from the median plane can be subject to associations covering the full median plane, rather than just the vertical dimension. Using 1/3-octave noise bands, Blauert [10] found 4 kHz and 16 kHz centered bands tended to be located at the front, 8 kHz from overhead, and 12 kHz from behind. Rogers and Butler [11] explained this effect in terms of 'covert peak areas' (CPAs) – which, for a frequency band, is the direction of a sound source of fixed distance from which maximum sound pressure is generated at the ear canal entrance. In the absence of other localization cues, subjects tend to localize such stimuli in the direction of its CPA. A relationship between CPA-localization and the pitch-height effect is not explicit.

1.2. The Volume Effect

The term 'volume' refers to the apparent size of an auditory image. This concept is also found in Stumpf's writings [1], and like other early theorists, he did not firmly distinguish volume from pitch-height phenomena. Many of the metaphors that can be drawn on to support pitch-height require little adjustment to apply to an association between intense low frequency sound and great size. Consider again 'chest' versus 'head' voice or woofer versus tweeter, as well as Zbikowski's pitch metaphor exceptions.

The study of volume and pitch-height diverged when investigated experimentally from the 1930s – volume studies being conducted using headphones, or a small loudspeaker behind the head, without regard for localization (or externalization). Stevens, whose key work in the development of volume used such presentation techniques, viewed it as not literally spatial [12]. Low frequency, intense, broadband, long duration, and low interaural cross-correlation stimuli are associated with large volume [13, 14]. Volume effects are found with tone and noise-band stimuli, and predictive models of volume have been proposed [13, 15].

In the field of auditorium acoustics, studies of spatial impression yield tendencies similar to those found in volume studies, most notably the effect of loudness and low frequency sound in expanding the auditory image [16-18] (although the research field emphasizes interaural characteristics). Such findings - in the context of externalized auditory space - suggest that the volume effect may be more than metaphorically spatial in audition.

The relationship between interaural characteristics and volume is easily understood in terms of auditory function, at least in general terms [19]. The present paper's Discussion considers whether there might be auditory bases for the influences of other determinants.

2. AIMS AND METHOD

The experiment described by this paper examined the two effects of height and image size for loudspeaker-sourced sound in an anechoic room. It aimed to see whether the volume effect would occur at all when referenced to externalized (visual) space, whether these effects can seperably coexist, and whether a vertically extended sound source is perceived as more extended than an equivalent compact source. This study also aimed to verify that the pitch-height effect occurs for noise-band stimuli, with the band spectrum and loudness as parameters.

Five two-way concentric loudspeakers (Tannoy System 800) were stacked on their sides in a vertical array, 0.12 m behind a lightweight black cloth screen marked with a square grid. The subjects sat in a height-adjustable chair with their ears 2 m from the center loudspeaker, which was at ear-height. Loudspeaker centers were at 0.28 m intervals, and so were elevated 0°, ±7.9° and ±15.6° with respect to ear height. Stimuli consisted of octave bands of noise, centered on 125 Hz, 500 Hz, 2 kHz and 8 kHz, as well as pink noise and filtered pink noise (lowpass at 3 kHz, –24 dB/oct.), presented in groups of ten 200 ms bursts (10 ms ramps). Two loudness levels were applied to the stimuli, following ISO532B [20], of 84 and 64 phons at the subject's head position. Stimuli were generated from either one, three or five contiguous loudspeakers, using every possible active array center. Where multiple loudspeakers were active, their signals were incoherent. The variation of distance and angle from the subject's head was accounted for with equalization and gain adjustments to each loudspeaker. Maximum time differences between loudspeakers were deemed too short to have a localization effect [21]. In summary, the independent variables were (i) signal spectrum, (ii) loudness level, (iii) number of active loudspeakers, and (iv) center loudspeaker of the active array — yielding 108 stimuli.

The grid (see Fig. 1) in front of the subjects had lines at 0.2 m intervals, identified with integers from –5 to +5. The grid origin was in line with the center loudspeaker (and hence the subject's head). Subjects were provided with a response sheet showing a scaled representation of the grid, and were instructed to mark the left, right, upper and lower limits of the apparent sound image. Subjects were unrestrained, but instructed to look at the grid center prior to trials. Subjects could also mark that the sound came from behind, or else that it was impossible to locate. The test involved eight training stimuli, and four sessions of 27 test stimuli.

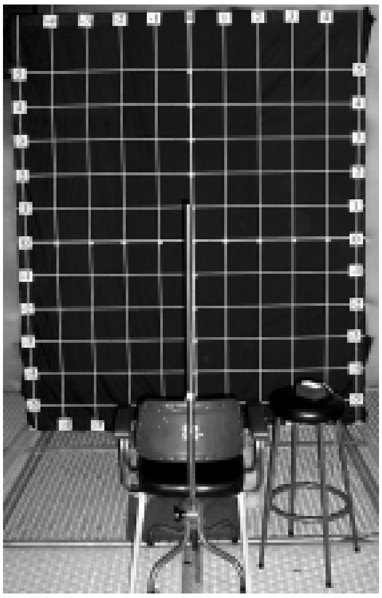

Figure 1. *The screen from behind the subject's chair.*

3. RESULTS

One subject's results were discarded because all stimuli were localized behind. For the 22 remaining subjects, 15% of judgements showed reversals and 4% could not be located.

Analysis of variance (ANOVA) for vertical image center (mean of the upper and lower image edges) found significant effects for signal spectrum and loudspeaker location ($F = 57.3$, $P < 0.0001$ and 46.3, <0.0001 respectively), with significant interaction (6.8, <0.0001), as well as significant interaction between signal spectrum and loudness level (6.8, <0.0001). Loudness level alone had a non-significant effect (2.7, 0.10). These effects are shown in Fig. 2.

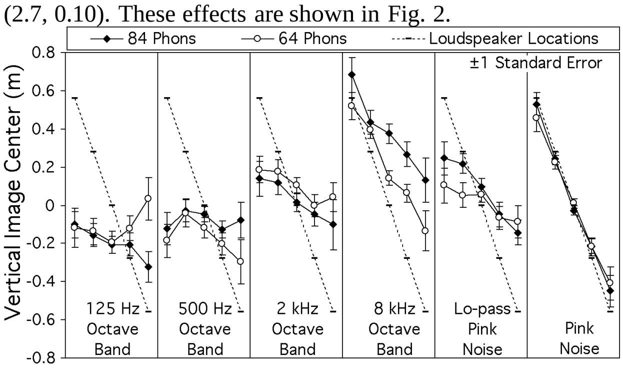

Figure 2. *Effects of signal spectrum, source location and loudness level on vertical image center.*

The source location appears to have at least some effect on the auditory image location for all signals, although the effect is marginal for those lacking high frequency energy. Localization is accurate for pink noise stimuli. With regard to the octave band signals, Scheffe tests show the pitch-height effect to be present for all comparisons except 125 Hz—500 Hz. Although the effect of loudness is modest, it does appear to enhance localization at 125 Hz, and have an elevating influence at 8 kHz.

The table below tallies significant results ($P < 0.05$) for Scheffe tests of each of the 22 subjects' vertical image center ratings for octave band stimuli. Results consistent with the pitch-height effect are followed by parenthesized contrary results.

	125 Hz	500 Hz	2 kHz
500 Hz	2 (1)	-	-
2 kHz	12 (2)	7 (0)	-
8 kHz	18 (0)	16 (0)	8 (0)

Horizontal and vertical spread were taken from the difference between left and right image edges, and between upper and lower edges respectively. ANOVAs found vertical and horizontal spread to be similarly affected by loudness level (vert. $F = 161.6$, $P < 0.0001$; horiz. 184.9, <0.0001), and more weakly by signal spectrum (vert. 4.2, <0.001; horiz. 20.8, <0.0001) – see Fig. 3. With one exception (8 kHz vs. LP noise, horizontal only), Scheffe tests only found significant differences between signal spectra for comparisons involving the 125 Hz band. There is an apparent tendency for low frequency stimuli to have greater horizontal than vertical spread, and high frequency stimuli to have greater vertical than horizontal spread.

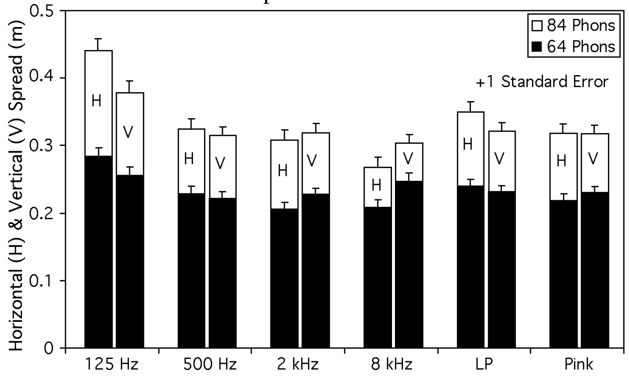

Figure 3. *Effects of loudness level and signal spectrum on horizontal and vertical image spread.*

Scheffe tests for individual subjects' image width judgements showed 19 (of 22) making significantly larger width judgements for the louder stimuli, and no significant contrary results. The Scheffe test image width tallies for the octave band stimuli are shown below:

	125 Hz	500 Hz	2 kHz
500 Hz	5 (0)	-	-
2 kHz	5 (0)	1 (0)	-
8 kHz	8 (0)	1 (0)	1 (0)

4. DISCUSSION

4.1. Key Findings

The experiment reproduced pitch-height and volume effects previously studied separately. It found the pitch-height effect present for octave-bands of noise, and the results for full-range and low-pass noise reiterate the importance of high frequency energy in accurate vertical localization.

The volume effect was revealed by both vertical and horizontal image expansion, and was unambiguously affected by the 20 phon loudness level interval. The effect of signal spectrum on volume was far more modest. The number of active loudspeakers in the vertical array did not significantly affect subjective ratings (including vertical spread).

While the volume effect for loudness level was evident in almost all individual subject results, the volume effect for signal spectrum was only evident for a few subjects. The pitch-height effect was evident, at least to some extent, for most subjects.

The volume effects were smaller than those observed in the traditional volume magnitude estimation experiments, as well as in auditorium sound-field experiments. For example, a 20 dB increase almost doubles the volume of a 1 kHz tone in Terrace and Stevens' model [13], and Perrott and Buell [14] find a 30 dB increase to yield an increase greater than six-fold in the volume of broadband noise. Auditorium studies show a 1.5°/dB increase in apparent source width [16], whereas the present study sees image width change from 9° to 13° with a ~20 dB increase. The use of loudspeakers and a visual grid in this experiment is the likely cause of its more modest results: the loudspeakers promote an externalized image, and the grid defines the spatial field explicitly, leaving little to the imagination.

4.2. Explanations of Effects

Although the pitch-height and volume effects found may be the result of response proclivity (where association influences responses in the absence of perceptual cues), the strength of results (including previous studies) makes explanations from auditory function attractive.

One plausible auditory explanation for the pitch-height effect may be construed by extending the notion of the head-related transfer function to include a reflective floor. Spectral cues for median plane localization are often studied in anechoic conditions, with features restricted to high frequencies mainly from pinna and shoulder reflections. However, on a reflective floor for a normal range of ear heights (from seated to standing), the spectral notches caused by interference between direct and reflected sound extend to low frequencies for elevated sources – meaning that a source on or near the floor will tend to convey more bass than an equivalent elevated source. It may be objected that walls and ceilings also make a large contribution to everyday auditory experience. Nevertheless, the floor (or ground) is almost always present, and at a height that is quite stable for an individual in a given posture, whereas wall distances and ceiling heights vary, and these features are often absent in the outdoors. For low frequencies, this extends the CPA-localization theory to the most general non-anechoic environment.

Whether the CPA-localization theory accounts for the effect at higher frequencies is more difficult to assess, because individuals' mappings of frequency to CPAs differ due to different pinna contours, and also because the CPA-localization effect is not restricted to the median plane's vertical dimension. Bearing in mind that the present experiment biased the subjects towards localizing the sound image on the screen in front of them, the tendency for narrow band signals around 8 kHz to have CPAs overhead might help explain the high elevation of the 8 kHz octave band in the present experiment [10]. This frontal bias is present in all experiments reporting the pitch-height effect, and absent in experiments directly investigating CPA effects.

Explaining the volume effect in terms of localization blur has some appeal, but at most it offers a partial explanation. The appeal is illustrated by visual analogy – an increase in the brightness of a blurred light source sees its apparent boundary expand. By analogy, as a sound becomes more intense, more of its perceptually blurred boundary is classified as part of the auditory image. However, the frequency region around 2 kHz (where horizontal localization blur is relatively high due to the cross-over between intensity and time cues) is not characterized by an increase in auditory image size. Furthermore, the larger vertical than horizontal localization blur is not reflected by larger vertical spread judgements in the present experiment.

Finally, the relationship between low frequency and volume is left unexplained.

The above hypothesis is refined by assuming that volume is a by-product of localization processes associated with interaural time differences. The interaural cross-correlation peak associated with a low frequency source is much broader than that for a high frequency source (because of the longer wave periods), and this sits well with the relationship between low frequencies and volume. It also provides a connection with studies in auditorium acoustics, where the height of the cross-correlation peak predicts apparent source width (a shorter peak makes a wider image).

4.3. Applications to Display

The fact that judgements of auditory size and height – made on an explicitly spatial visual grid – were affected by the non-spatial parameters of loudness and center frequency suggests a robustness in these parameters suitable for display applications. Although the experiment was conducted in anechoic conditions, both the pitch-height and volume effects have been previously found in non-anechoic conditions [4,6,7,16,17].

Almost all subjects show the pitch-height effect (at least, for extreme frequency contrasts) and almost all associate size with loudness. Display design must bear in mind that a small proportion of users may be insensitive to these effects.

Compared with loudness, the predominant frequency of noise bands only weakly affected image size. In a broader study on auditory display parameters, Walker [22] found that when representing size with frequency, a majority of subjects chooses low frequency to represent large size, but a significant number chooses the reverse. Hence it seems better to use loudness than frequency in the auditory display of size, despite the frequency-volume relationship in the auditory volume literature. Loudness was used in this way by Evreinov [23] to represent size in an auditory display.

Neuhoff et al [24] sound a note of caution in using pitch in auditory display. They show large differences between the performance of subjects with and without musical training in identifying the size and direction of frequency intervals, also noting the incidence of tone-deafness. It must be appreciated that the frequency ranges of the present and previous pitch-height effect studies were much larger than the melodically-plausible range studied by Neuhoff et al.

5. CONCLUSIONS

This study contrasts effective auditory spatial cues from signals (loudness and source spectrum) with the potential for physical spatial parameters (source vertical extent and center) to be ineffective. It reaffirms that, without head-related transfer function manipulations, auditory stimuli can have varying spatial attributes in addition to environmentally-generated spatial cues. Such attributes should lend support to the inexpensive use of sound for spatial representation, especially when in harmony with extrinsic associations and widely-used metaphors.

6. REFERENCES

[1] C. Stumpf, *Tonpsychologie*, S. Hirzel Leipzig, Germany, 1883 and 1890 (2 volumes).

[2] L.M. Zbikowski, “Metaphor and music theory: reflections from cognitive science,” *Music Theory Online*, vol. 4, no. 1, http://societymusictheory.org/mto/, Jan. 1988.

[3] R. Scruton, “Understanding music,” *Ratio*, vol. 25, no. 2, pp. 97-120, 1983.

[4] C.C. Pratt, “The spatial character of high and low tones,” *J. Exp. Psych.*, vol. 13, pp. 278-285, 1936.

[5] F.L. Dimmick and E. Gaylord, “The dependence of auditory localization upon pitch,” *J. Exp. Psych.*, vol. 17, pp. 593-599,1934.

[6] O. Trimble, “Localization of sound in the anterior-posterior and vertical dimensions of auditory space,” *Brit. J. Psych.*, vol. 24, pp. 320-334, 1934.

[7] S.K. Roffler and R.A. Butler, “Location of tonal stimuli in the vertical plane,” *J. Acoust. Soc. Am.*, vol. 43, pp. 1260-1266, 1968.

[8] S.K. Roffler and R.A. Butler, “Factors that influence the localization of sound in the vertical plane,” *J. Acoust. Soc. Am.*, vol. 43, pp. 1255-1259, 1968.

[9] V.R. Algazi, C. Avendano and R.O. Duda, “Elevation localization and head-related transfer function analysis at low frequencies,” *J. Acoust. Soc. Am.*, vol. 109, no. 3, pp. 1110–1122, Mar. 2001.

[10] J. Blauert, “Sound localization in the median plane,” *Acoustica*, vol. 22, pp. 205-213, 1969.

[11] M.E. Rogers and R.A. Butler, “The linkage between stimulus frequency and covert peak areas as it relates to monaural localization,” *Perception and Psychophysics*, vol. 52, no. 5, pp. 536-546, 1992.

[12] S.S. Stevens, “Are tones spatial?” *Am. J. Psych.*, vol. 46, pp. 145-147, 1934.

[13] H.S. Terrace and S.S. Stevens, “The quantification of tonal volume,” *Am. J. Psych.*, vol. 75, pp. 596-604, 1962.

[14] D.R. Perrott and T.N. Buell, “Judgements of sound volume: effects of signal duration, level and interaural characteristics on the perceived extensity of broadband noise,” *J. Acoust. Soc. Am.*, vol. 75, pp. 1413-1417, 1982.

[15] D. Cabrera, “The size of sound: auditory volume reassessed,” *Mikropolyphonie*, 5, farben.latrobe.edu.au/mikropol/, 1999.

[16] R. Williamson, “Optimisation of variable lateral energy for spatial impression in a hall,” *Appl. Acoust.*, vol. 26, pp. 113-134, 1989.

[17] M. Morimoto and Z. Maekawa, “Effects of low frequency components on auditory spaciousness,” *Acustica*, vol. 66, pp. 190-196 1988.

[18] D. Cabrera, “Auditory image size: Volume and spatial impression,” in *Proc. 8th Western Pacific Acoustics Conf.*, Melbourne Australia, 2003.

[19] A. Marshall and M. Barron, “Spatial responsiveness in concert halls and the origins of spatial impression,” *Appl. Acoust.*, vol. 62, pp. 91-108, 2001.

[20] E. Zwicker, H. Fastl and Dallmayr, “BASIC program for calculating the loudness of sounds from their 1/3 octave band spectra according to ISO 532 B,” *Acustica*, vol. 55, pp. 63, 1984.

[21] J. Blauert, “Localization and the law of the first wavefront in the median plane,” *J. Acoust. Soc. Am.*, vol. 50, pp. 466-470, 1971.

[22] B. Walker, *Magnitude Estimation of Conceptual Data Dimensions for use in Sonification*, PhD, Rice University, 2000.

[23] G. Evreinov, “Spotty: imaging sonification based on spot-mapping and tonal volume,” in *Proc. ICAD*, Helsinki Finland, pp. 67-72, 2001.

[24] J.G. Neuhoff, R. Knight and J. Wayand, “Pitch change, sonification, and musical expertise: Which way is up?” in *Proc. ICAD*, Kyoto Japan, pp. 351-355, 2002.