Role of spectral detail in sound-source localization

Abhijit Kulkarni & H. Steven Colburn

Hearing Research Center and Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, USA

Sounds heard over headphones are typically perceived inside the head[1] (internalized), unlike real sound sources which are perceived outside the head (externalized). If the acoustical waveforms from a real sound source are reproduced precisely using headphones, auditory images are appropriately externalized and localized[1–4]. The filtering (relative boosting, attenuation and delaying of component frequencies) of a sound by the head and outer ear provides information about the location of a sound source by means of the differences in the frequency spectra between the ears as well as the overall spectral shape. This location-dependent filtering is explicitly described by the head-related transfer function (HRTF) from sound source to ear canal. Here we present sounds to subjects through open-canal tube-phones and investigate how accurately the HRTFs must be reproduced to achieve true three-dimensional perception of auditory signals in anechoic space. Listeners attempted to discriminate between 'real' sounds presented from a loudspeaker and 'virtual' sounds presented over tube-phones. Our results show that the HRTFs can be smoothed significantly in frequency without affecting the perceived location of a sound. Listeners cannot distinguish real from virtual sources until the HRTF has lost most of its detailed variation in frequency, at which time the perceived elevation of the image is the reported cue.

Although previous studies[1–4] showed that real and virtual sounds (matched to have identical acoustic waveforms in the ear canals) were indistinguishable when the subject was wearing headphones, they did not address the question of the sensitivity of the subjects to the details of the HRTF. Here we presented virtual sounds to subjects' ears through open-canal 'tube-phones' (Fig. 1), so that 'real' sounds were essentially unaffected by the presence of the tube-phones and sounds from real and virtual sources could be directly compared. The paired-comparison experiments required listeners simply to report the order of the stimuli: real first or virtual first. Subjects were given practice and trial-by-trial feedback so that small differences in the location or the realism of the sounds could be used for judgements. In the measurements with smoothed HRTFs, the stimulus spectrum was randomized so that non-spatial attributes (such as timbre) could not be used for discrimination.

In the first experiment, four subjects were tested for their ability to distinguish natural stimuli from virtual stimuli, which were constructed to match the natural stimuli exactly. Source spectra were not varied (beyond the stochastic nature of the noise waveforms) in these initial validation experiments. We tested four azimuthal locations (0, 45, 135 and 180 degrees, where 0 degrees is straight ahead) in separate sets of trials. None of the subjects was able to distinguish natural from virtual stimuli. Performance of each subject at each location was within the bounds expected from chance performance. (The percentages of correct values obtained are plotted with the data from the smoothing experiment in Fig. 2.) These results were consistent with the subjects' impressions that they perceived both types of stimuli as completely natural and were unable to perceive any differences between the virtual and the free-field stimuli. These basic results show the adequacy of our methods for measurements of HRTFs and for virtual-stimulus presentation as well as the naturalness of the resulting perceptions.

In a second set of experiments, we found that, provided that the overall interaural time delay (ITD) was maintained to be consistent with that in the natural sound waveforms, virtual sounds synthesized through HRTFs with the original magnitude spectra (that is, that of the natural stimulus) but simplified phase spectra were also indistinguishable from the natural sounds. This lack of dependence of the spatial percept on the frequency-dependent detail of the ITD is consistent with previous results[1,5,6] and led us to develop our procedure in which the modified HRTFs were given the minimum-phase response[7] for each magnitude, supplemented by a pure ITD that was calculated to make the ITD of the virtual stimulus match the ITD of the real stimulus.

In all of our other experiments, the magnitude spectra of empirical HRTFs measured from individual listeners were systematically smoothed and the dicriminability of the resulting virtual stimuli from free-field stimuli was assessed. This smoothing was performed by expressing the HRTF log-magnitude spectrum as a Fourier series and reconstructing the HRTF spectrum using a truncated series (see Methods). The nature of the smoothing operation is shown in Fig. 3 for a representative HRTF for six values of the smoothing parameter. For these experiments, the signal components of the test stimuli (the source waveforms) were

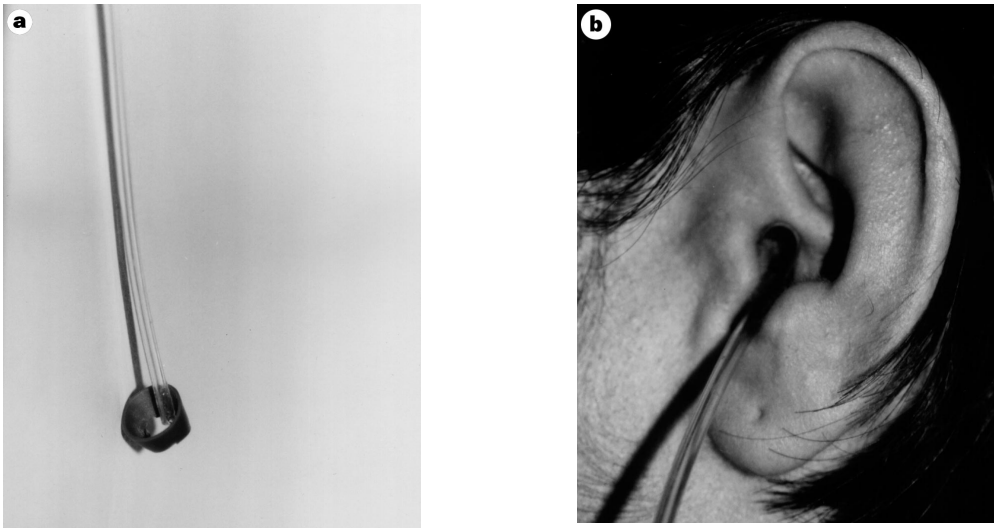

Figure 1 The tube-phone apparatus. **a**, The apparatus attached to a customized latex-rubber hollow shell to fit the ear canals of listeners. **b**, The apparatus inserted in the ear canal of a listener. Note that the shell provides a snug fit flush with the walls of the ear canal with minimal distortion of ear-canal acoustics.

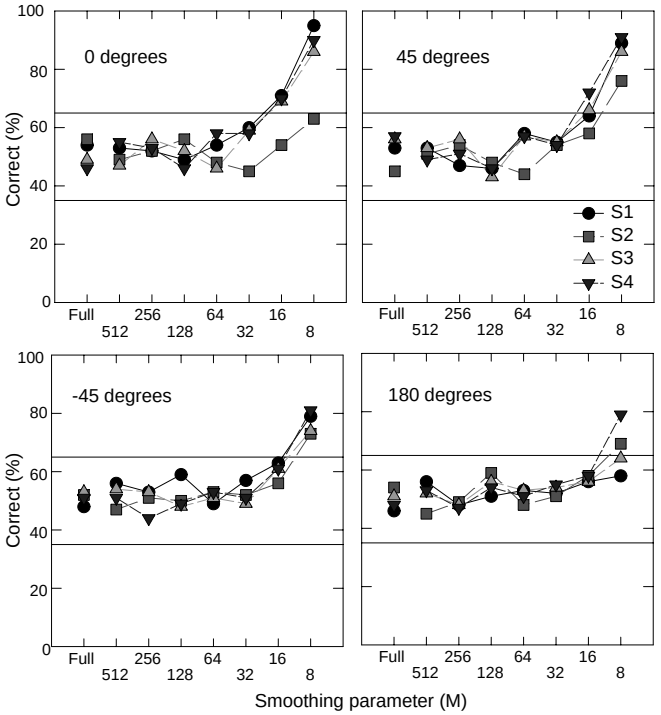

Figure 2 Discrimination performance of four subjects (S1–S4) as a function of the smoothing parameter, *M* (see Methods), at each azimuthal position tested. The 95% confidence bounds for chance performance are shown by the two horizontal lines in each graph. The left-most points in each panel (marked 'Full' along the abscissa) correspond to performance in response to the complete HRTF and no stimulus randomization.

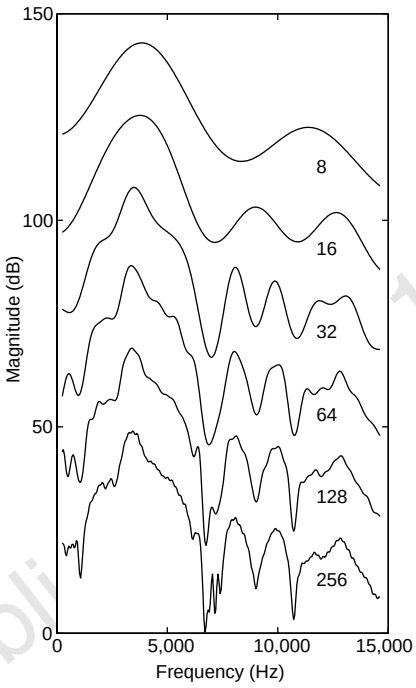

Figure 3 Spectral smoothing of the left-ear HRTF magnitude spectra measured from a representative subject (S1) for the 0-degree location. The HRTFs are shown on a relative scale and are offset from each other by 20 dB. The smoothing parameter, *M*, is described in the Methods and is indicated next to each curve.

randomly varied in each trial. The interval-to-interval variation (see Methods) in the stimulus waveform introduced an uncertainty in the quality of the sounds in each presentation and prevented listeners from using non-spatial differences in sound quality (such as timbre) to discriminate between natural and virtual auditory images. We determined a priori that this variation did not alter the location of the image from a loudspeaker.

The performance of the four subjects tested is shown in Fig. 2 for four azimuthal locations (0, 45, 135 and 180 degrees). The performance (in per cent correct) is plotted as a function of the smoothing parameter (the number of Fourier coefficients used in the HRTF reconstruction). Performance was largely unaffected by the spectral smoothing; performance was always within the bounds of chance performance for reconstructions having more than 16 coefficients. In fact, performance was seldom outside these bounds except at the extreme smoothing condition (8 Fourier coefficients). Comparisons of Figs 2 and 3 show that discrimination performance remained within chance bounds for fairly large amounts of smoothing and became consistent only at the extreme smoothing conditions tested.

The subjective reports of all listeners were consistent with the results reported above. All subjects reported hearing both free-field and virtual stimuli from the free-field loudspeaker in almost all trials. It was more surprising, however, that, even at the extreme smoothing condition (8 Fourier coefficients) when the discrimination performance of subjects was high, all subjects reported complete externalization of the virtual sound image. In fact, all subjects reported that the only useful cue in the extreme smoothing conditions was the elevation of the virtual image (to varying degrees) above the free-field sound image. The elevation of the virtual image in the extreme smoothing conditions is consistent with the observation that the magnitude spectra of HRTFs from high elevations are relatively smooth compared with those from lower elevations. The compelling perception of a well-externalized sound was also perceived by several experienced psycho-acousticians using spectrally smooth stimuli presented through the tube-phone apparatus. All of our observations are consistent with the conclusion that spectral detail is not important for sound externalization and localization.

Our results show that, at least in anechoic space when no head movements occur and when the visual cues are consistent and accurate, details of the HRTFs (magnitude and phase) are not responsible for a well-localized, externalized perception of a sound source. These conclusions do not indicate that individual variations in the HRTF are unimportant, as we used individualized HRTFs in all of our experiments. Our results are consistent with studies in which crude approximations to the natural ear-input signals were perceived as natural provided that these waveforms were made to change in a manner consistent with the movement of the listener's head[8]. We suggest that a critical factor contributing to the success of our experimental technique is the playback apparatus: the tube-phone system preserves the acoustics of a natural listening condition. Specifically, both the pathways for internally and externally generated sounds and the ear-canal impedances are consistent with the natural (open-ear) listening condition. In contrast to suggestions of previous studies[1,2,4] that quality deteriorates when HRTFs are not accurately reproduced, our results show that the fine structure of the HRTF is relatively unimportant for auditory spatial attributes, including externalization. □

Methods

Tube-phone apparatus. Virtual field stimuli were presented through the ER-2 tube-phone (Etymotic Research) which was used without the provided foam-tips. The use of the tube-phones without foam-tips results in a second-order (40 dB per decade), low-frequency (below 2 kHz) roll-off in the frequency response; we compensated for this roll-off by applying a second-order pre-emphasis filter over the relevant frequency region. The virtual acoustical signals

were computer-generated using individualized HRTFs measured from the entrance of the blocked ear canal (for reasons of repeatability and good signal-to-noise ratio[9]). A position-independent transformation was then applied to these blocked-canal HRTFs to make the pressure at the eardrum of the listener during playback consistent with the free-field pressure at that point. Natural-sound field stimuli were presented from a loudspeaker which was located 1.15 m from the listener's head. Both the real and the natural stimuli had a bandwidth of 350–15,000 Hz.

The presence of the 1.35-mm-diameter tubes in the ear canals had negligible effects on the natural sound field. This was observed subjectively, and confirmed by acoustical measurements and by a study comparing localization performance in the median-saggital plane with and without the tube-phones present. We also measured the acoustical crosstalk between the tube-phones at the two ears to be negligible at the presentation levels used (50 dB sound pressure level (SPL) measured with steady-state excitation in the ear canal of the KEMAR acoustical mannequin).

Psychophysical discrimination paradigm. The experimental paradigm was a two-interval, two-alternative, forced-choice (2I, 2AFC) discrimination task. Each interval consisted of a pair of 80-ms noise bursts (having 10-ms raised-cosine rise/fall times) presented at 45 dB SPL and separated by 100 ms. The time between the two intervals of each trial was 200 ms. The stimuli in each interval were either synthesized virtually or delivered from a free-field loudspeaker in an anechoic chamber (from the corresponding location), in random order. In the experiments in which the HRTF spectrum was smoothed, the signal components of the noise stimuli were randomized in 1/3-octave bands by ± 5 dB in each trial to prevent listeners from using any non-spatial differences in sound quality in discriminating the virtual sounds from the real sounds. In a localization experiment conducted in the median-saggital plane we determined that the localization performance of listeners was not significantly different with or without the spectral rove in the noise stimulus.

The location of a single source was tested during each run and feedback regarding the correct answer was provided in each trial. HRTFs were measured from listeners at the start of each experimental run with head position monitored by a Polhemus head-position sensor (which was worn on top of the subject's head, held by a plastic frame). A computer program instructed the subject to find 0-degree yaw, pitch and roll coordinates; the program read the Polhemus sensor output and delivered voiced instructions (consisting of digitized speech samples) over a loudspeaker. The instructions consisted of the words 'turn', 'roll', 'left', 'right', 'up', 'down', and 'hold it'. The measurement was only made when the reference position had been reached. The subject remained seated in the same location for the rest of the session and was prompted (if necessary) to achieve the same orientation for each presentation during playback. Head position was constantly monitored during the course of the stimulus playback and the trial was automatically terminated if any motion was detected. This ensured that discrimination between the stimuli did not result from any misalignment between the listener and the speaker and that no dynamical cues were used to determine the source of the stimulation.

HRTF smoothing. The discrete Fourier transform, $H(e^{j\omega})$, of an HRTF impulse response can be expressed as a discrete sequence, $H[k]$, where $H[k] = H(e^{j\omega})|\omega = (2\pi/N)k$, where N is the number of sample values. The discrete realization of the log-magnitude spectrum may be expressed by the Fourier series:

$$\hat{H}[k] = \log|H[k]| = \sum_{n=0}^{N/2} C(n)\cos(2\pi nk/N)$$

where $C(n)$ is the nth coefficient in the Fourier series for the sequence $\log|H[k]|$.

By using a partial Fourier series of degree $M < N/2$, a smoothed fit to the HRTF spectrum can be obtained:

$$\hat{H}[k] = \sum_{n=0}^{M} C(n)\cos(2\pi nk/N)$$

The choice of $M = N/2$ corresponds to an exact reconstruction of the HRTF magnitude spectrum and the choice of $M = 1$ corresponds to a reconstruction having a flat spectrum at the average value of the HRTF.

For an empirical HRTF impulse response having N sample values, a Fourier series having $M = N/2$ terms provides an exact reconstruction of the empirical magnitude spectrum. The empirical HRTF in our experiments consisted of 1,024 time points (and thus correspond to a Fourier series with 512 terms). The smoothed magnitude spectra used in our experiments were constructed using 256, 128, 64, 32, 16 and 8 Fourier coefficients. The HRTFs were implemented as minimum-phase filters, augmented with a frequency-independent time delay, which was consistent with the overall ITD in the empirical HRTF measurements for the location being tested.

Received 10 August; accepted 23 October 1998.

1. Hartmann, W. M. & Wittenberg, A. On the externalization of sound images. *J. Acoust. Soc. Am.* **99,** 3678–3688 (1996).
2. Wenzel, E. M. Localization in virtual acoustic displays. *Presence Teleop. Virt. Environ.* **1,** 80–107 (1992).
3. Wightman, F. L. & Kistler, D. J. Headphone stimulation of free-field listening I: stimulus synthesis. *J. Acoust. Soc. Am.* **85,** 858–867 (1989).
4. Zahorik, P., Wightman, F. L. & Kistler, D. J. in *Proc. Acoustics, Speech and Signal Processing (Institute of Electrical and Electronics Engineers) Workshop Appl. Signal Processing Audio Acoustics* (IEEE Press, New York, 1995).
5. Kistler, D. J. & Wightman, F. L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.* **91,** 1637–1647 (1991).
6. Kulkarni, A., Isabelle, S. K. & Colburn, H. S. Sensitivity of human subjects to head-related transfer function phase spectra. *J. Acoust. Soc. Am.* (submitted).
7. Oppenheim, A. V. & Schafer, R. W. *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
8. Loomis, J. M. & Soule, J. I. in *Proc. Society for Information Display 1996 Int. Symp.* 965–968 (Society for Information Display, San Jose, CA, 1996).
9. Shaw, E. in *Binaural and Spatial Hearing in Real and Virtual Environments* (eds Gilkey, R. H. & Anderson, T. R.) 25–48 (Lawrence Erlbaum, Mahwah, NJ, 1997).

Correspondence and requests for materials should be addressed to A.K. (e-mail: abhijit@enga.bu.edu).

Hedgehog stimulates maturation of Cubitus interruptus into a labile transcriptional activator

Johanna Talavera Ohlmeyer & Daniel Kalderon

Department of Biological Sciences, Columbia University, 1212 Amsterdam Avenue, New York, New York 10027, USA

In *Drosophila*, signalling by the protein Hedgehog (Hh) alters the activity of the transcription factor Cubitus interruptus (Ci) by inhibiting the proteolysis of full-length Ci (Ci-155) to its shortened Ci-75 form[1,2]. Ci-75 is found largely in the nucleus and is thought to be a transcriptional repressor[1], whereas there is evidence[3–5] to indicate that Ci-155 may be a transcriptional activator[1,2,6]. However, Ci-155 is detected only in the cytoplasm, where it is associated with the protein kinase Fused (Fu), with Suppressor of Fused (Su(fu)), and with the microtubule-binding protein Costal-2 (refs 1,7–9). It is not clear how Ci-155 might become a nuclear activator. We show here that mutations in *Su(fu)* cause an increase in the expression of Hh-target genes in a dose-dependent manner while simultaneously reducing Ci-155 concentration by some mechanism other than proteolysis to Ci-75. Conversely, eliminating Fu kinase activity reduces Hh-target gene expression while increasing Ci-155 concentration. We propose that Fu kinase activity is required for Hh to stimulate the maturation of Ci-155 into a short-lived nuclear transcriptional activator and that Su(fu) opposes this maturation step through a stoichiometric interaction with Ci-155.

Hh secreted from posterior compartment cells of the wing disc in *Drosophila* induces expression of *decapentaplegic* (*dpp*) in anterior cells as far as the site of the future vein 3 (refs 10, 11) and anterior *engrailed* (*en*) expression in a narrower strip of cells close to the anterior–posterior (AP) compartment border[11,12] (Fig. 1a). The amount of Dpp protein secreted from the AP border dictates the position at which vein 2 forms[13]. En and other factors induced by Hh contribute to the formation of vein 3 and associated campaniform sensillae at the anterior edge of the Hh-signalling territory[14].