Supporting Information

Genzel et al. 10.1073/pnas.1712058115

A direct comparison of motion profiles for the AM (green) and PM (blue) conditions is shown in Fig. S1. This subject performed the active motion quite precisely and the motion platform could reproduce the motion reliably. This pertains both to the (dominating) azimuthal motion (Fig. S1*A*) and the associated small vertical motion (Fig. S1*B*).

SI Materials and Methods

Reasons for Band-Pass Filtering of the Stimuli in Exp. I. The reason for restricting the frequency range to 800–4,000 Hz is that at the low end (below 800 Hz), we have to avoid near-field effects, some of them being dependent on absolute frequency, not only on the relationship between frequency and sound-source diameter. The reason for the low-pass cutoff at 4 kHz is that for higher frequencies we would have stronger occlusion effects of the farther speaker by the closer speaker (see below). Nevertheless we needed enough bandwidth to accommodate several harmonics to provide a reasonably strong pitch. Due to the dominance region of pitch (1), it is likely that the high-pitched pip train had higher pitch strength than the low-pitched train.

Motion Training and Body Motion Analysis for Exp. I. At the beginning of the first experimental session, each subject was taught to move in a stereotypical way for the AM condition. They had to start moving their upper body such that their head was displaced 23 cm to the left within the first second of stimulus playback, then 46 cm to the right (for a target displacement of 23 cm to the right from the head origin) within the following 2 s, and back to the starting point within the final second, in an overall smooth motion akin to a single period of a sine wave (Fig. S1). This motion profile was trained before every session with feedback from the experimenter, who instructed the subjects to move, started the playback of a 4-s click train similar to the stimulus used in the experiment, and immediately analyzed the head tracking data. This procedure was repeated until the subject was confident that they had memorized the motion profile and the experimenter observed several subsequent trials with acceptable head tracks (correct velocity of motion, displacement to the left and right between 20 and 26 cm, and stable position along the other two spatial axes).

After the session, the head tracks for each trial were analyzed as to whether they met the inclusion criteria: NM trials were excluded when any tracking point acquired during stimulus presentation deviated from the head origin by more than 2 cm along the interaural axis, or when the mean absolute deviation from that point exceeded 1 cm for the trial as a whole. AM trials were excluded when the maximum displacement to the left or to the right along the interaural axis differed by more than 4 cm from the mean displacement for the subject.

Calibrations for Exp. I. To calibrate setup with real sound sources, a measurement microphone (1/2″; BSWA Technology) was positioned at the point corresponding to the middle of the interaural axis of a seated subject. For each pair of loudspeaker depths, acoustic impulse responses of the speakers were measured and compensation impulse responses calculated by pointwise division of the complex discrete Fourier transform (DFT) of an ideal band-pass IR between 200 and 8,000 Hz by the complex DFT of the measured IR. All sounds presented through the two speakers were convolved with the corresponding compensation impulse responses. This procedure equalized loudness, spectrum, and latency differences between the two speakers.

To remove possible residual spectral or loudness cues that may contribute to distance discrimination, we implemented a roving spectral envelope. Specifically, we defined a random spectral envelope by varying loudness across a ±6-dB range in third-octave steps throughout the whole calibrated pass band of the speakers (200–8,000 Hz). Thus, the timbre of the harmonic complexes changed from trial to trial which renders the use of timbre or near-field cues very difficult. The validity of this precise equalization procedure plus the application of the spectral-envelope rove was psychophysically confirmed by the fact that our subjects performed poorly when they were not allowed to move during stimulus presentation.

To check for residual spectral effects that may arise through interaction of the sound sources with the subject's head or torso, we made control measurements replacing the subject with a head and torso simulator (B&K 4128C).

Head Tracking in Exp. I. Tracking was implemented with a camera on the subject's head scanning a target made up of fiducial markers mounted at the ceiling above the subject (2). Stimulus presentation for each trial was started only when the head position did not vary by more than 1 cm along the interaural axis, 1.5 cm along the anterior–posterior axis, or 1.5 cm along the cranial–caudal axis from the required head origin. The head origin was defined at the beginning of each experimental session as that head position where the distance from the interaural axis to the membrane of the front loudspeaker at its closest position was exactly 30 cm, and the head was exactly on axis with the two loudspeakers.

Rendering of Virtual Sound Sources on the Loudspeaker Array in Exp. II (Amplitude Panning Procedure). For each of the two virtual sound sources, the horizontal axis of the speaker array was intersected with the line between the sound source and the most recently acquired head position of the subject. The two speakers closest to this intersection point were activated to simultaneously reproduce the respective sound source. The ratio of their gains was chosen according to the distance of the center points of those loudspeakers to the intersection point: When one activated speaker was at a distance a and the other at a distance b from the intersection point, their gains were set according to the ratio b:a. The combined gain of the two speakers was set to account for geometric attenuation due to the distance between the subject's head position and the virtual sound source. Loudspeaker activations and gain settings were updated at a rate of 100 Hz throughout stimulus playback.

Motion Training and Body Motion Analysis for Exp. II. To ensure comparability of the results between the three conditions that involved motion of the sound sources relative to the subjects' heads (AM, PM, and SSM), both motion training and inclusion criteria for motion trials were more rigorous than in Exp. I. All subjects underwent precise training concerning the active body motion that they had to perform in the AM condition. Small markers on the speaker array indicated the leftmost, rightmost, and middle positions the subjects had to meet in this sequence during their motion. The experimenter informed the subject during training if the motion matched the targeted motion profile.

During the main data acquisition, trials were excluded when they did not meet a nested set of criteria that quantified deviations of the executed motion in that trial from the targeted motion profile. These trials were repeated again at a later time until at least 30 trials per condition were obtained.

When subjects had learned to reliably reproduce body motion with the required displacement and velocity, a further training procedure was initiated. Here subjects were moved by the platform or conducted their learned body motion, but additionally the sound sources were presented with the largest source-distance difference and the subjects had to decide whether the high-pitched source was closer or farther away than the low-pitched source. One training block consisted of 120 trials. This training was necessary because with virtual sound sources and the many different interleaved conditions, it was somewhat harder for the subjects to exploit auditory motion parallax. The main experiment could begin only after a subject's performance in a training block was at least 80% correct.

1. Ritsma RJ (1967) Frequencies dominant in the perception of the pitch of complex sounds. *J Acoust Soc Am* 42:191–198.
2. Garrido-Jurado S, Muñoz-Salinas R, Madrid-Cuevas FJ, Marín-Jiménez MJ (2014) Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognit* 47:2280–2292.

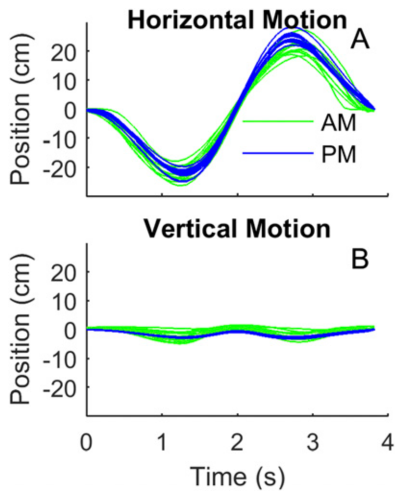

Fig. S1. Tracks of horizontal (*A*) and vertical (*B*) head motion of the subject relative to the stationary sound sources when the subject moved either actively (green) or the subject was moved by the motion platform (blue). Data show that subjects were successfully trained to move quite stereotypically and that the platform captured this stereotypical motion quite well.