Psychophysical evidence for auditory motion parallax

Daria Genzel[a,b,1], Michael Schutte[a,1], W. Owen Brimijoin[c], Paul R. MacNeilage[b,d,2], and Lutz Wiegrebe[a,b,3]

[a]Department Biology II, Ludwig Maximilians University Munich, 82152 Planegg-Martinsried, Germany; [b]Bernstein Center for Computational Neuroscience Munich, 82152 Planegg-Martinsried, Germany; [c]Glasgow Royal Infirmary, Medical Research Council/Chief Scientist Office Institute of Hearing Research (Scottish Section), G31 2ER Glasgow, United Kingdom; and [d]Deutsches Schwindel- und Gleichgewichtszentrum, University Hospital of Munich, 81377 Munich, Germany

Edited by Wilson S. Geisler, University of Texas at Austin, Austin, TX, and approved February 16, 2018 (received for review July 6, 2017)

Distance is important: From an ecological perspective, knowledge about the distance to either prey or predator is vital. However, the distance of an unknown sound source is particularly difficult to assess, especially in anechoic environments. In vision, changes in perspective resulting from observer motion produce a reliable, consistent, and unambiguous impression of depth known as motion parallax. Here we demonstrate with formal psychophysics that humans can exploit auditory motion parallax, i.e., the change in the dynamic binaural cues elicited by self-motion, to assess the relative depths of two sound sources. Our data show that sensitivity to relative depth is best when subjects move actively; performance deteriorates when subjects are moved by a motion platform or when the sound sources themselves move. This is true even though the dynamic binaural cues elicited by these three types of motion are identical. Our data demonstrate a perceptual strategy to segregate intermittent sound sources in depth and highlight the tight interaction between self-motion and binaural processing that allows assessment of the spatial layout of complex acoustic scenes.

depth perception | distance discrimination | spatial hearing | self-motion | auditory updating

Humans' dominant sense for space is vision. The exceptional spatial resolution and acuity of foveal–retinal vision allows for accurate and simultaneous localization of multiple objects in azimuth and elevation (1). The observer's distance to an object, however, is more difficult to assess. In vision, distance of near objects is mainly encoded by binocular disparity which relies on image differences resulting from the spatially separate views of the two eyes onto the object (2–4); these differences become minimal for far objects. Higher visual centers integrate disparity with information arising from monocular cues, many of which provide information about relative depth separation, rather than absolute distance. These include occlusion of one object by another one, relative size, perspective, shading, texture gradients, and blur (5, 6). Important information about relative depth is also added when there is motion of the observer relative to the environment or object: the resulting difference in image motion between features at different depths is termed motion parallax (3, 6). In the case of observer motion, relative depth from motion parallax can be scaled to obtain absolute estimates of object distance if information about speed of observer motion is available, for example, based on vestibular signals (7). Such scaling cues are generally not available when the object moves relative to the stationary observer.

Apart from the visual system, only audition allows locating objects (i.e., sound sources) in the far field beyond the range of touch. As in vision, azimuth and elevation of the sound sources are readily encoded through auditory computation, both binaural (interaural time and level differences, ref. 8) and monaural (elevation-dependent analysis of pinna-induced spectral interference patterns, ref. 9). But again, the distance to a sound source is most difficult to assess: in the absence of reverberation, and without a priori knowledge about the level and spectral composition of the emitted sounds, distance estimation for humans is indeed impossible (10). This is not surprising, considering that an important visual distance cue (binocular disparity) is not available in audition, not least because humans cannot point each of their ears toward a sound source. Some visual depth cues have auditory counterparts, (e.g., blur is related to frequency-dependent atmospheric attenuation, and relative size to loudness), but many others are unavailable (e.g., occlusion, texture gradients, shading).

In reverberant rooms, the ratio of the sound energy in the first wave front relative to the energy reflected from the surfaces is a function of distance and allows the estimation of sound-source distance without motion (11–14). Recent theoretical work has pointed out that motion of the interaural axis (and specifically translational head motion) also allows fixing sound-source distance, through the analysis of auditory motion parallax (15). To date, however, it is unexplored to what extent auditory motion parallax may be exploited by human subjects to perceptually segregate sound sources in distance and how the time-variant binaural cues that are generated by translational head motion are integrated with vestibular and/or proprioceptive cues for auditory distance perception. After an early report that "head movement does not facilitate perception of the distance of a source of sound" (16) work by Loomis and coworkers (17, 18)

Significance

When we cannot see a sound source, it is very difficult to estimate how far away it is. In vision, motion parallax facilitates depth perception in that when we move, nearer objects move more in our visual field than farther objects. Our experiments show that humans can also exploit motion parallax for sound sources. Moreover, we show that, as in the visual system, facilitation of auditory depth perception is stronger when the subjects move actively than when subjects are being moved or when only the sound sources move. We conclude that dedicated mechanisms exist that integrate self-motion with binaural auditory motion cues, allowing humans to exploit auditory motion parallax.

Author contributions: D.G., M.S., W.O.B., P.R.M., and L.W. designed research; D.G., M.S., and P.R.M. performed research; D.G., M.S., P.R.M., and L.W. analyzed data; and D.G., M.S., W.O.B., P.R.M., and L.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Data have been deposited in the G-Node Network of the Bernstein Center for Computational Neuroscience (available at https://doid.gin.g-node.org/5103a1db3d918f82a1724c8d90f6ca0b/).

See Commentary on page 3998.

[1]D.G. and M.S. contributed equally to this work.

[2]Present address: Department of Psychology, Cognitive and Brain Sciences, University of Nevada, Reno, NV 89523.

[3]To whom correspondence should be addressed. Email: lutzw@lmu.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1712058115/-/DCSupplemental.

Published online March 12, 2018.

has shown that dynamic binaural cues elicited by translational self-motion relative to a stationary sound source may provide some (rather erroneous) information about the absolute distance of a sound source for tested source distances between two and six meters. More recent work has highlighted the interaction of self-motion (real or visually induced) on the perception of auditory space: Teramoto et al. (19, 20) have shown that self-motion distorts auditory space in that space is contracted into the direction of self-motion, regardless of whether the self-motion was real (which provided a vestibular signal) or visually induced (which provides no vestibular input but only visually mediated self-motion information). However, it remains unclear whether self-motion can support the segregation of sound sources in distance through an auditory motion parallax and how proprioceptive and vestibular inputs may contribute to this segregation.

Here we present formal psychophysical data showing that humans can segregate a high-pitched sound source from a low-pitched sound source in distance based on the time-variant binaural perceptual cues associated with motion. The initial demonstration of auditory motion parallax is implemented as a forced-choice experiment with real sound sources positioned at different depths in anechoic space that have been carefully calibrated to eliminate nonmotion-based cues to distance. In a second experiment, we instead elicit differences in perceived depth of sound sources positioned at the same depth by rendering sounds contingent on head tracking, and we show that this exploitation of auditory motion parallax is facilitated by both vestibular and proprioceptive information arising from active self-motion.

Results

Seven subjects were asked to respond whether a high-pitched sound source was closer or farther away than a low-pitched source. The two sound sources were temporally interleaved, i.e., the sum of the sound sources was perceived as alternating in pitch over time at a rate of 10 Hz for each source or 20 Hz for the summed sources (*Materials and Methods*).

Careful steps were taken to eliminate nonmotion-based auditory cues to distance. Consequently, at each position of the sound sources, sound level and spectral content of each sound source was identical when measured either with an omnidirectional microphone at the center of the subjects' interaural axis or when measured binaurally with a Bruel & Kjaer (B&K) 4128C Head-and-Torso Simulator (*Materials and Methods*). An illustration of the experimental setup, the stimulus, and the psychophysical results is shown in Fig. 1.

When the sound sources were separated in distance by only 16 cm (leftmost data in Fig. 1*C*), subjects could not solve the task when they were not allowed to move; performance was around chance level, 50% (black symbols and line). However, when the subjects were allowed to move their heads laterally by ±23 cm (green symbols and line), performance was much better and subjects scored on average 75% correctly even at the smallest presented distance difference of 16 cm. With increasing distance difference between the sound sources, performance quickly improved when active self-motion was allowed while performance stayed rather poor without active motion. Nevertheless, some subjects could discriminate the sound sources without self-motion for larger distance differences. Possible residual distance cues are discussed below. Of the 42 pairs of performances (seven subjects times six distance differences) performance in the active-motion condition (AM) was significantly better than in the no-motion condition (NM) in 29 cases. In no case was performance better without motion than with motion (Fisher's exact test, $P < 0.05$, P values corrected for multiple testing with the Benjamini–Hochberg procedure). These data clearly show that human subjects can easily exploit auditory motion parallax to segregate sound sources in depth. To this end, subjects likely exploit time-variant binaural cues arising from the lateral self-motion: with a given lateral motion, the closer object creates larger binaural cues because it covers a larger range of azimuthal angles relative to the subject's moving head. The role of self-motion and its interaction with dynamic binaural processing is further explored in the following experiment.

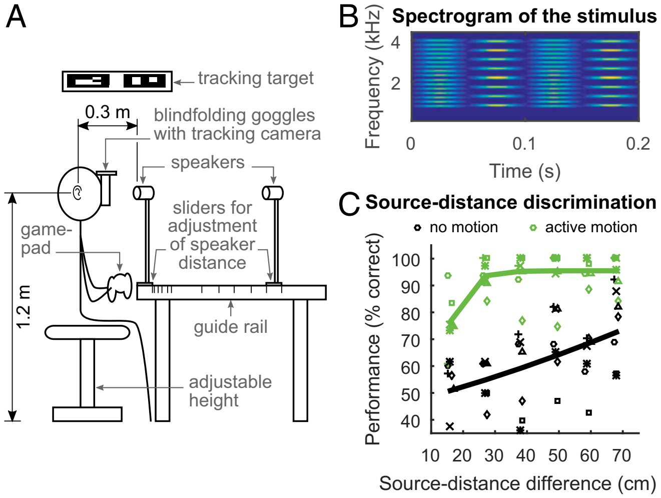

Fig. 1. Illustration of the experimental setup (*A*), the stimuli (*B*), and psychophysical results (*C*) to demonstrate auditory motion parallax in Exp. I. (*A*) Subjects were seated with their interaural axis exactly perpendicular to the axis of two miniature broadband loudspeakers. One randomly chosen speaker emitted the high-pitched sound, the other speaker emitted the low-pitched sound. Head motion in each trial was continuously recorded with a head-tracking system consisting of a tracking camera on the subjects' heads and a tracking target at the ceiling. (*B*) Spectrogram of a 0.2-s section of the intermittent low- and high-pitched stimulation in each trial. The two different pitches are presented by the two speakers at different depths. (*C*) Individual performances (marked by different symbols) and sigmoidal fit to average performance (solid lines) with motion (green) and without motion (black). The data show that with motion, subjects discriminate sound-source distances overall quite well, whereas performance hardly deviated from chance level without motion.

Here sound sources were presented in virtual space via a linear high-resolution loudspeaker array which precluded the use of distance-dependent loudness and reverberation cues, so that it was not necessary to change calibration dependent on virtual-source distance (*Materials and Methods*). The motion conditions were as follows: NM, subjects remained positioned with their head in line with the two sound sources at different distances; AM (Fig. 2 *A* and *B*), subjects actively moved their upper body by about 23 cm left and right following a previously trained motion profile (these two conditions were the same as in the first experiment) (*Materials and Methods*); passive motion (PM) (Fig. 2*C*): subjects did not move, but the subjects were moved by a motion platform such that the subject's head moved in the same way as in the AM condition; and sound-source motion (SSM) (Fig. 2*D*), subjects remained still but the sound sources presented via the array moved such that the relative motion between the sound sources and the subject's head in azimuth was the same as in the AM and PM conditions. Twelve subjects took part in this second experiment.

Without any motion of either the sound sources or the subjects, none of the subjects could reliably determine whether the high-pitched sound source was nearer or farther than the low-pitched sound source. Performance of an example subject in the experimental condition without motion (NM) is represented by the black asterisk in Fig. 3. The failure to discriminate distances is not surprising because loudness cues related to absolute distance were quantitatively removed and the use of the speaker array for virtualization (*Materials and Methods*) precluded the

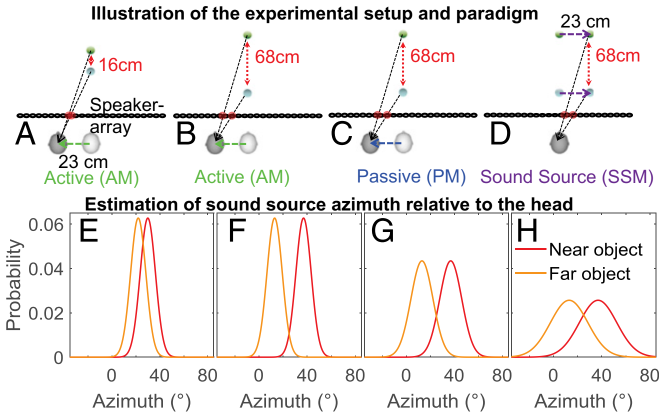

Fig. 2. Illustration of the setup and paradigm of Exp. II and the hypothesis. (*A*–*D*) Subjects were trained to move parallel to the speaker array with the same motion profile as in Exp. I. Subjects performed the motion either actively (AM) (*A* and *B*), or they were moved by a motion platform (PM) (*C*). In these conditions, tracking of the head motion relative to the array and the virtual sound sources allowed us to update the speaker activation in real time. In the SSM condition (*D*) the sound sources moved along the array but the subjects were stationary. Speaker activation is illustrated by the red area around the speakers. (*E*–*H*) With increasing depth separation, dynamic binaural cues get stronger (*E* and *F*). AM provides additional information (proprioceptive and efference copy signal) and leads to better discrimination (*F*). During PM, only vestibular signals provide additional information (*G*). Discrimination is therefore worse than for AM, but better than for SSM, where only dynamic binaural cues are present (*H*).

use of differential reverberation cues for distance estimation. When we trained our subjects to move laterally during the presentation of the alternating high- and low-pitched sources, the subjects improved their ability to identify which sound source was nearer. In principle, this question can be answered by identifying the nearer source as the one whose perceived azimuthal angle changes more during the lateral self-motion. An example of depth discrimination performance as a function of source distance difference is shown in Fig. 3. Performance in the AM condition is shown in green. This subject could reliably judge whether the high-pitched source was closer or farther away than the low-pitched source when the closer source was 40 cm and the farther source was 56 cm away from the subject, i.e., the distance difference was only 16 cm. However, performance deteriorated when the subject was passively moved by a motion platform (PM, blue curve), or when the sound sources moved (SSM, purple curve).

The validity of the direct comparison between the motion conditions depends on the precision of the actively executed motion and how well this motion is reproduced by the motion platform. A comparison of the active and passive motion profiles is found in *Supporting Information*.

Distance-difference thresholds are shown in Fig. 4*A*, individual data represented by the colored bars and Fig. 4*B*, medians and interquartiles represented by the box plots. The data clearly show that subjects performed best [just-noticeable distance differences (JNDs) were smallest] when they actively moved in front of the virtual sound sources (AM). Performance was significantly worse when subjects were moved by the motion platform (PM). When the subjects were stationary but the sound sources moved (SSM), thresholds were worst. In this condition, some of the subjects could not solve the task even for the largest source-distance difference, 68 cm. In Fig. 4, data from these subjects are artificially set to a threshold of 80 cm; note, however, that real perceptual thresholds may be larger. In summary, these data confirm that also with virtual sound sources, subjects can resolve distance differences between sound sources quite well when they move in a manner that exploits auditory motion parallax. The fact that they performed worse with passive motion indicates that both proprioceptive and vestibular signals are integrated with dynamic binaural cues to solve the task. Visual cues were unavailable because the subjects were blindfolded. Without proprioceptive and vestibular signals, i.e., without motion of the subject, performance was significantly worse, which shows that the dynamic binaural cues alone (which were the same in all three conditions of Exp. II) do not suffice to provide the best performance. Results in the AM condition compare well across Exps. I and II: The average threshold for 75% correct performance was about 16 cm sound-source difference in Exp. I and 20 cm source difference in Exp. II. This is true although the setups differed substantially (real sound sources in Exp. I vs. simulated sound sources in Exp. II).

Discussion

The current psychophysical experiments support the hypothesis that human subjects can exploit auditory motion parallax to discriminate distances of sound sources. Thus, the capacity to exploit motion parallax to disambiguate sensory scenes is shared between the senses of vision and audition. Importantly, subjects received no trial-to-trial feedback about their performance in Exp. I. When asked to respond to whether the high-pitched sound source was closer or farther than the low-pitched source, subjects appeared to readily exploit motion parallax when they were instructed to move, without extensive training. The perceptual basis for auditory motion parallax is that, through lateral motion of either the objects or the subject, the distance difference between the objects is transferred into time-variant horizontal localization cues. For a given lateral motion, the closer object produces the stronger variation in horizontal localization cues, i.e., interaural time differences (for the lower part of the

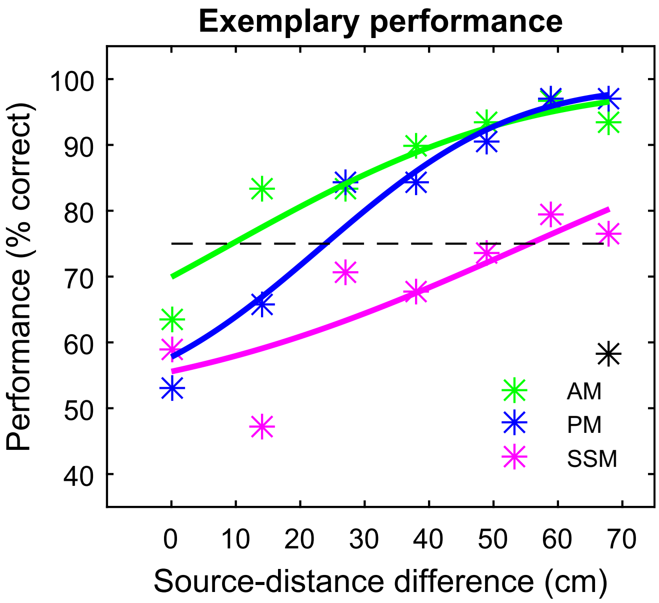

Fig. 3. Exemplary performance (symbols) and fitted psychometric functions (lines) for depth discrimination of two alternating sound sources as a function of their distance difference in Exp. II. Performance is best in the AM condition (green) where the subject performed an active head motion and worse in the SSM condition (purple) where the subject was stationary, but the sound sources moved past him or her. When the subject was moved by the motion platform past the virtual sound sources (PM, blue), performance was intermediate. Without both subject- and sound-source motion (NM), the subject could not solve the task even at the largest distance difference of 68 cm (single black star). Therefore, full psychometric functions were not obtained in the NM condition.

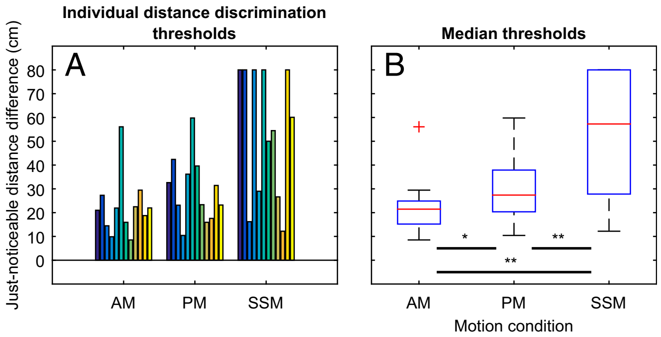

Fig. 4. Psychophysical performance thresholds (just-noticeable differences) for sound-source distance in Exp. II. Individual data are shown by the colored bars in *A*; boxplots of medians (red) and interquartiles (blue boxes) are provided in *B*. Whiskers represent the data range expressed as the 75th percentile plus 1.5 times the difference between the 75th and the 25th percentile (maximum range) and the 25th percentile minus 1.5 times the difference between the 75th and the 25th percentile (minimum range). The red cross in *B* represents the only value outside the whisker range (outlier). Nonparametric paired comparisons (Wilcoxon signed rank tests) show that performance in the AM condition is significantly better than in both the PM (*$P < 0.05$, signed rank = 10) and the SSM condition (**$P < 0.01$, signed rank = 4), and that performance in the PM condition is significantly better than in the SSM condition (**$P < 0.01$, signed rank = 6).

stimulus spectrum below about 1 kHz) and level differences (for the higher-frequency parts) change faster for the closer sound source than for the farther source. Thus, while self-motion may have limited value in estimating absolute distance to a single sound source (17, 18), the current experiments demonstrate that self-motion readily supports segregation of sound sources in depth.

The dynamic binaural cues in the current experimental conditions with motion (AM, PM, and SSM) are equally salient, no matter whether the subject moves actively, the passive subject is moved, or the objects move. Nevertheless, the current data show that subjects are more sensitive to distance differences when they move actively than when they are moved or when the objects move.

For the visual system, it has long been known that viewers can use motion parallax to estimate distances of objects (2, 3, 21) not only by humans but also by, e.g., Mongolian gerbils (22). Interestingly, also in vision, distance estimation is better when the viewer moves than when the objects move (23). Thus, the current data corroborate previous conclusions, drawn for the visual system, that self-motion information facilitates the depth segmentation of sensory scenes.

While with virtual sound sources (Exp. II) subjects failed completely to discriminate sound-source distances without motion (cf. Fig. 3), some subjects could discriminate large distance differences between real sound sources (Exp. I, data in black in Fig. 1*C*). Close inspection of binaural room impulse responses recorded from the two sound sources with a head-and-torso simulator indicate that this may be related to residual low-frequency reflections in the experimental booth. The booth was fully lined with acoustic foam of 10-cm thickness, resulting in a lower cutoff frequency of the damping to around 1 kHz. Given that the lower cutoff of our stimulation was at 800 Hz, it is possible that some subjects exploited residual reverberation cues to solve the distance discrimination task even without motion. Nevertheless, the data clearly show that motion-induced perceptual cues are dominant in solving the task.

In purely geometric terms, there is a limit to the extent to which motion parallax may be used to resolve a difference in distance between two sources. Assuming perfect detection, quantification, and temporal integration of an observer's own physical motion, successful source-distance discrimination could only occur if the subject's motion were to result in a difference in subtended angle between the two sources that is equal to or larger than the minimum detectable change in source angle over time. In the auditory system, this limit is imposed by the minimum audible movement angle; in the visual system, it is imposed by the spatial displacement threshold. The fundamental constraint applied by these angular acuity thresholds may be formalized in Eq. **1**:

$$d' = \tan\left(a\tan\left(\frac{d}{x}\right) - \Theta\right) \times x, \qquad [1]$$

where d is the distance of the farther target, x is the amount of lateral motion, Θ is the angular acuity threshold, and d' is the distance to a closer target that is just discriminable.

At ideal source velocities, signal characteristics, and contrasts, the lowest auditory motion detection threshold is roughly 2° (24, 25), whereas the threshold in the visual system is at least 100 times smaller at roughly 1 arcmin, or 0.017° (26). In the framework of Eq. **1** it is clear that the visual system should be more capable of using motion parallax to discriminate distance than the auditory system. By using each modality's values for Θ in Eq. **1**, we can estimate that for a maximum lateral displacement of 23 cm from the loudspeaker axis and a distance of the farther target of 52 cm, the auditory system should begin to detect a difference when the closer target was at about 47 cm. In Exp. II, only our best subject could reliably discriminate 45 cm from 52 cm, i.e., a distance difference of 7 cm in the AM condition. Thus, even with optimal cue combination, our subjects performed worse than predicted from auditory motion detection of a single sound source.

In the visual system, on the other hand, in an equivalent task with the same lateral motion, a difference should become perceivable with the closer object being only 4 mm closer than the farther object. In practice, parallax distance acuity in the visual system may be yet more accurate even than this, due to the ability to compare signals at the two eyes (27) and the use of eye motion itself (28), a mechanism unavailable to the human auditory system. Critically, parallax-based distance discrimination becomes poorer as a function of distance for both visual and auditory objects. Given the lower spatial acuity, this is especially impactful for auditory signals: for a sound source at 4 m and an orthogonal listener motion of 20 cm, a second sound source would have to be about 1.6 m closer to be discriminable in depth.

These computations assume a perfect assessment and use of observer motion. Combination of motion signals with other sensory input is known to be imperfect and this has been established in the visual system (29, 30), auditory system (31), and even the somatosensory system (32). Given this additional source of error, it is likely that the true depth discrimination thresholds are higher than estimated by Eq. **1**. Larger physical motion would necessarily increase distance acuity, however, and the motion limits used here may not accurately reflect natural behavior, particularly for a walking individual.

The current results are in line with previous work showing that dynamic binaural processing works best under the assumption that sound sources are fixed in world coordinates and dynamic binaural changes are assumed to be generated by self-motion: Brimijoin and Akeroyd (33) measured minimum moving audible angles (MMAAs), i.e., the minimum perceivable angle between two (speech) sound sources when both sounds rotated relative to the subject's head. The authors showed that the MMAA was significantly smaller when the subject's head rotated but the sound sources were kept fixed in world coordinates than when the head was kept fixed and the sound sources were rotated around the subject. As in the current study, the authors took care that dynamic binaural cues were the same in the two experimental conditions.

Given the accumulating evidence suggesting that binaural processing, and even auditory distance computation, is facilitated by self-motion, it is important to consider how this facilitation takes place: we assume that vestibular and proprioceptive cues (and of course visual cues, if available) allow the generation of a prediction about the velocity and position of auditory targets. This prediction acts as additional information, which according to standard cue-integration models (34), leads to a reduced variance in the combined estimate. This argumentation is illustrated in the *Lower* panels of Fig. 2, referenced to the experimental conditions illustrated in the respective *Upper* panels: In the SSM condition (Fig. 2 *D* and *H*), the lack of nonauditory cues results in an imprecise representation of the azimuth of the two sound sources. The distributions overlap significantly, and depth discrimination based on these representations will be poor. In the PM condition (Fig. 2 *C* and *G*), vestibular information is integrated with the auditory information, leading to a decrease in the variance of the representations. In the AM condition (Fig. 2 *B* and *F*), proprioceptive information is also integrated, resulting in a further decrease of the variance. This decrease in variance allows for more reliable discrimination and consequently better thresholds (Fig. 2 *A* and *E*). Overall we argue that auditory motion parallax is a classical illustration of how a combination of cues from different modalities supports object-discrimination performance.

It could be expected then, that passive self-motion leads to less facilitation, and exclusive sound-source motion removes all nonauditory cues. It was suggested that the ratio of motion to visual pursuit encodes depth information from motion parallax better than motion or pursuit alone (35). In the current auditory study, distance discrimination also improved when the subjects were actively moving, and this improvement might be a result of a similar ratio of self-motion to binaural auditory pursuit. The fact that our subjects performed significantly better when they moved actively than when they were moved or when the sound sources moved supports this hypothesis because the motion is less well defined when it lacks the proprioceptive component (passive vs. active motion) and explicit motion information is missing completely when only the sources move (SSM). In the latter condition, subjects are likely to fall back on the use of pursuit information alone and consequently perform still worse. Overall the good correspondence between the current results and those on visual motion parallax support the hypothesis that the current experiments may tap into a dedicated multimodal motion parallax circuit.

Regardless of the exact nature of the underlying circuit, we conclude that distance discrimination in the current study was based solely on parallax cues. While recent studies indicate that the classical binaural cues (interaural time and/or level differences) also depend on distance, at least when the sound source is in the near field, i.e., quite close to the subject (36, 37), these effects cannot account for the present results. Even though the positions of the virtual sound sources were quite close to the subjects (between 30 and 98 cm), near-field effects can be excluded because the loudspeakers used to present the virtual sound sources were very small (membrane diameter of <2.5 cm) and frequencies relatively high (≥800 Hz). With these parameters, the near field extends to no more than 6 cm in front of the array, even when two adjacent speakers are active at a time (38).

Where would such a "sensitivity" to acoustic distance cues be computed in the brain? A possible candidate for neuronal representation of auditory distance might be the auditory "where" pathway. Indeed Kopčo et al. (39) found that the posterior superior temporal gyrus and planum temporal were activated by the above-mentioned auditory distance cues like the direct-to-reverberant ratio (11) and distance-dependent interaural level differences (36). It would be very promising to include active or passive motion into such scanning paradigms and test the extent to which motion enhances neural activity in the spatial–auditory brain areas, however challenging this might be for brain-imaging techniques.

Materials and Methods

The current psychophysical experiments were approved by the Ethics Committee of the Ludwig Maximilians University Munich, project no. 115–10. All subjects signed an informed consent protocol.

Exp. I.

Stimuli. Subjects were required to judge the relative distances of two intermittent sound sources. Each of the sources emitted a train of tone pips with a pip duration of 25 ms and a repetition period of 100 ms. The carrier for the tone pips was a harmonic complex with a fundamental frequency of either 210 Hz (low-pitched source) or 440 Hz (high-pitched source). For reasons detailed in *Supporting Information*, pips were band-pass filtered to cover the same frequency range between 800 and 4,000 Hz, i.e., the fundamental and (at least for f0 = 210 Hz) a few lower harmonics were missing in both sources. The phase of the low-pitched pip train was shifted by 50 ms, relative to the high-pitched train, such that the overall stimulation consisted of a summary pip train with a 50-ms period and periodically alternating pitch. A spectrogram of the summary pip train with alternating pitches is shown in Fig. 1*B*.

In Exp. I, the pips were played back through two miniature speakers (NSW1-205–8A, AuraSound) positioned at different depths in front of the subject in an anechoic chamber. The speakers were mounted on vertical rods that were fitted to mechanical sliders moving in a guide rail (see Fig. 1*A*). This construction allowed the speakers to be precisely positioned in depth while minimizing the mutual acoustic shadowing of the speakers. Relative to the subjects' interaural axis, the source distances were (at increasing level of difficulty) 98/30 cm, 90/31 cm, 82/33 cm, 73/35 cm, 65/38 cm, and 56/40 cm. This resulted in distance differences between the sound sources of 68, 59, 49, 38, 27, and 16 cm. Without the spectral rove (see below), the sound level of the pip trains was 67 dB sound pressure level. The loudspeakers were driven via a stereo amplifier (Pioneer A107) from a PC soundcard. In each trial, the closer loudspeaker pseudorandomly emitted either the low-pitched or the high-pitched pip train, and the more distant loudspeaker emitted the other pip train. Detailed information on our acoustic calibrations is provided in *Supporting Information*.

Procedure. In a one-interval, two-alternative forced choice paradigm with feedback, subjects had to judge whether the high-pitched sound source was closer or farther away from them than the low-pitched source by pressing one of two buttons on a gamepad. The subjects were seated throughout the experiment. Their heads were continuously tracked. Head tracking procedures are detailed in *Supporting Information*. At the beginning of each trial, a 100 ms pure-tone burst at 1 kHz informed the subjects when their head had reached an acceptable position. Then subjects were instructed to either remain in that position during the following 4-s stimulus presentation (for the NM condition) or to make a trained ±23-cm lateral motion (for the AM condition) (*Supporting Information*, *Motion Training and Body Motion Analysis for Exp. I*).

Within each block of 20 trials (10 NM trials and 10 AM trials), loudspeaker positions were fixed. Data were acquired in at least four sessions of six blocks each. Trials were included or excluded based on the respective head tracks (see below), and data acquisition was continued until at least 30 acceptable trials were available for each experimental condition and pair of loudspeaker depths. Reported distance-difference thresholds correspond to the 75% correct value extracted from a cumulative Gaussian distribution fitted to the data.

Subjects. Seven female subjects (ages ranging from 22 to 38 y) participated in the experiment. None of the subjects reported auditory, vestibular, or sensory–motor impairments.

Exp. II. Stimuli, procedure, and data analysis for Exp. II were the same as for Exp. I with the following exceptions:

Stimuli. In contrast to Exp. I, the sound sources were presented with a loudspeaker array which allowed positioning the sound sources in virtual space behind the array. The array consisted of 24 miniature broadband speakers (NSW1-205–8A, AuraSound) spaced at a distance of 4 cm. Each speaker was individually equalized with a 64-point finite impulse response filter to provide a flat magnitude and phase response between 200 Hz and 10 kHz.

The sound presentation was controlled via SoundMexPro (HörTech GmbH) allowing for dynamically adjusting the loudness of each speaker during playback. Sounds were sent out by a multichannel audio interface (MOTU 424 with two HD192 converters, MOTU, Inc.) and amplified with four multichannel amplifiers (AVR 445, Harman Kardon).

Procedure. The subject's head and the motion platform on which the subject was seated (see below) were continuously tracked with a 6-degree-of-freedom tracking system (Optitrack Flex 13, three cameras; NaturalPoint) sampling at 120 frames per second. The readings from the tracking system were used during stimulation to map the virtual sound sources to the speaker array by means of an amplitude panning procedure (for details see *Supporting Information*).

Exp. II was conducted on a 6-degree-of-freedom motion platform (Moog 6DOF2000E). Blindfolded subjects were seated in a padded seat mounted on the platform. All experiments were performed in a darkened room. The PC also controlled the platform. The tracking system sent its acquired data to a second PC. Both computers were connected via Ethernet.

Subjects initiated each trial by positioning their heads facing the middle of the speaker array (between loudspeaker 12 and 13) at a distance to the array of 20 cm. The press of a gamepad button started the presentation of the two sound sources via the speaker array. While the sound sources were on, the subjects or the sound sources moved, depending on the instructed motion condition, which included the two conditions studied in Exp. I (NM and AM) plus two additional conditions. In the PM condition (Fig. 2*C*), subjects did not move their upper body, but the platform moved the subjects such that the subjects' head motion relative to the virtual sound sources was very similar to the trained motion in the AM condition. In the SSM condition (Fig. 2*D*), subjects remained positioned with their heads directed toward the middle speakers but the sound sources presented via the array moved such that the relative motion between the sound sources and the subjects' heads in azimuth was the same as in the AM and PM conditions.

In contrast to Exp. I, the subjects received auditory feedback after every trial, indicating whether their decision was correct or not.

For each of the three motion conditions AM, PM, and SSM, 210 trials were collected per subject, 30 repetitions for each of the seven source-distance differences. A total of 90 additional trials were collected for the NM condition, but only at the largest distance difference of 68 cm. The overall 720 trials were divided into six blocks of 120 trials each. Trials for all conditions and sound-source distances were presented in a predefined randomly interleaved sequence in a given experimental block. Subjects were instructed about what kind of motion was required for the next trial. See *Supporting Information*, *Motion Training and Body Motion Analysis for Exp. II* for details.

Subjects. Twelve subjects, four males and eight females (ages ranging from 21 to 37 y), participated in the experiment. Two of the subjects also took part in Exp. I. None of the subjects reported auditory, vestibular, or sensory–motor impairments.

ACKNOWLEDGMENTS. We thank Benedikt Grothe and the Bernstein Center in Munich for providing excellent research infrastructure. We thank Isabelle Ripp for her assistance in data acquisition. This work was supported by a grant (01GQ1004A) of the Bernstein Center Munich, BMBF (Federal Ministry of Education and Research, Germany), and the Munich Center for Neuroscience.

1. Westheimer G (2005) The resolving power of the eye. *Vision Res* 45:945–947.
2. Rogers B, Graham M (1979) Motion parallax as an independent cue for depth perception. *Perception* 8:125–134.
3. Helmholtz H (1925) *Helmholtz's Treatise on Physiological Optics* (Optical Society of America, New York).
4. Qian N (1997) Binocular disparity and the perception of depth. *Neuron* 18:359–368.
5. Howard IP, Rogers BJ (2002) *Seeing in Depth*, Depth Perception (Porteus, Toronto), Vol 2.
6. Howard IP, Rogers BJ (1995) *Binocular Vision and Stereopsis* (Oxford Univ Press, New York).
7. Dokka K, MacNeilage PR, DeAngelis GC, Angelaki DE (2011) Estimating distance during self-motion: A role for visual-vestibular interactions. *J Vis* 11:2.
8. Rayleigh JWS (1879) XXXI. Investigations in optics, with special reference to the spectroscope. *London Edinburgh Dublin Philos Mag J Sci* 8:261–274.
9. Blauert J (1997) *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).
10. Coleman PD (1962) Failure to localize the source distance of an unfamiliar sound. *J Acoust Soc Am* 34:345–346.
11. Bronkhorst AW, Houtgast T (1999) Auditory distance perception in rooms. *Nature* 397:517–520.
12. Zahorik P (2002) Direct-to-reverberant energy ratio sensitivity. *J Acoust Soc Am* 112: 2110–2117.
13. Kolarik AJ, Moore BC, Zahorik P, Cirstea S, Pardhan S (2016) Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss. *Atten Percept Psychophys* 78:373–395.
14. Bekesy GV (1938) Über die Entstehung der Entfernungsempfindung beim Hören. *Akust Z* 3:21–31.
15. Kneip L, Baumann C (2008) Binaural model for artificial spatial sound localization based on interaural time delays and movements of the interaural axis. *J Acoust Soc Am* 124:3108–3119.
16. Simpson WE, Stanton LD (1973) Head movement does not facilitate perception of the distance of a source of sound. *Am J Psychol* 86:151–159.
17. Speigle JM, Loomis JM (1993) Auditory distance perception by translating observers. *Proceedings of 1993 IEEE Research Properties in Virtual Reality Symposium* (San Jose, CA), pp 92–99.
18. Loomis JM, Klatzky RL, Philbeck JW, Golledge RG (1998) Assessing auditory distance perception using perceptually directed action. *Percept Psychophys* 60:966–980.
19. Teramoto W, Sakamoto S, Furune F, Gyoba J, Suzuki Y (2012) Compression of auditory space during forward self-motion. *PLoS One* 7:e39402.
20. Teramoto W, Cui Z, Sakamoto S, Gyoba J (2014) Distortion of auditory space during visually induced self-motion in depth. *Front Psychol* 5:848.
21. Wexler M, van Boxtel JJ (2005) Depth perception by the active observer. *Trends Cogn Sci* 9:431–438.
22. Ellard CG, Goodale MA, Timney B (1984) Distance estimation in the Mongolian gerbil: The role of dynamic depth cues. *Behav Brain Res* 14:29–39.
23. Panerai F, Cornilleau-Pérès V, Droulez J (2002) Contribution of extraretinal signals to the scaling of object distance during self-motion. *Percept Psychophys* 64:717–731.
24. Saberi K, Perrott DR (1990) Minimum audible movement angles as a function of sound source trajectory. *J Acoust Soc Am* 88:2639–2644.
25. Strybel TZ, Manligas CL, Perrott DR (1992) Minimum audible movement angle as a function of the azimuth and elevation of the source. *Hum Factors* 34:267–275.
26. Lappin JS, Tadin D, Nyquist JB, Corn AL (2009) Spatial and temporal limits of motion perception across variations in speed, eccentricity, and low vision. *J Vis* 9:1–14.
27. McKee SP, Taylor DG (2010) The precision of binocular and monocular depth judgments in natural settings. *J Vis* 10:5.
28. Naji JJ, Freeman TC (2004) Perceiving depth order during pursuit eye movement. *Vision Res* 44:3025–3034.
29. Furman M, Gur M (2012) And yet it moves: Perceptual illusions and neural mechanisms of pursuit compensation during smooth pursuit eye movements. *Neurosci Biobehav Rev* 36:143–151.
30. Freeman TC, Champion RA, Warren PA (2010) A Bayesian model of perceived head-centered velocity during smooth pursuit eye movement. *Curr Biol* 20:757–762.
31. Freeman TC, Culling JF, Akeroyd MA, Brimijoin WO (2017) Auditory compensation for head rotation is incomplete. *J Exp Psychol Hum Percept Perform* 43:371–380.
32. Moscatelli A, Hayward V, Wexler M, Ernst MO (2015) Illusory tactile motion perception: An analog of the visual Filehne illusion. *Sci Rep* 5:14584.
33. Brimijoin WO, Akeroyd MA (2014) The moving minimum audible angle is smaller during self motion than during source motion. *Front Neurosci* 8:273.
34. Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415:429–433.
35. Nawrot M, Stroyan K (2009) The motion/pursuit law for visual depth perception from motion parallax. *Vision Res* 49:1969–1978.
36. Kuwada CA, Bishop B, Kuwada S, Kim DO (2010) Acoustic recordings in human ear canals to sounds at different locations. *Otolaryngol Head Neck Surg* 142:615–617.
37. Kim DO, Bishop B, Kuwada S (2010) Acoustic cues for sound source distance and azimuth in rabbits, a racquetball and a rigid spherical model. *J Assoc Res Otolaryngol* 11:541–557.
38. Weinzierl S (2008) *Handbuch der Audiotechnik* (Springer, Berlin).
39. Kopčo N, et al. (2012) Neuronal representations of distance in human auditory cortex. *Proc Natl Acad Sci USA* 109:11019–11024.