Supporting Information

Kopčo et al. 10.1073/pnas.1119496109

SI Results

Additional Analysis of the Behavioral Experiment Data. The main goal of the behavioral experiment was to confirm that the listeners could use the intensity-independent cues to judge source distance in our simulated auditory environment. The virtual space simulation had limitations due to technical constrains of the fMRI-compatible stimulus presentation equipment and due to the fact that nonindividualized binaural room impulse responses (BRIRs) were used in the simulation (1). Fig. 2 shows that, on average, the subjects could judge distance well, despite these limitations. To evaluate individual subjects' performance, the sensitivity index d' was estimated for each subject on the basis of his/her discrimination performance. The estimation was based on a psychophysical decision theory model (2) that assumed that (*i*) the listener's internal percept evoked in response to a stimulus presented from distance s can be described by a Gaussian-distributed random variable X; (*ii*) the mean value of X grows logarithmically with the actual source distance s; (*iii*) the variance in the internal representation, corresponding to internal noise, is fixed for each listener, independent of the actual stimulus distance, presentation intensity, or any other stimulus parameter; and (*iv*) each listeners' discrimination responses are unbiased optimal decisions based solely on the currently observed two values of X, corresponding to the two source distances presented in a given trial. On the basis of these assumptions, the following equation defines the percenage of correct performance, P_C, as a function of d' and the locations of the stimulus sources:

$$P_c = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{d'}{2}} e^{\frac{-t^2}{2}} \mathrm{d}t,$$

where $d' = |\ln s_1 - \ln s_2|/\sigma$. The constant σ [with the units of ln (cm)] was fitted for each subject to minimize the mean squared error between the subject's average performance for each source pair and the predicted performance based on the subject's estimate of σ. On the basis of these estimates, individual subject performance prediction was computed for each distance interval from Fig. 2. Gray line in Fig. 2 plots the across-subject averages based on this performance estimate. There is a good agreement between the gray and black lines in Fig. 2, confirming that the assumptions of the model and the individual estimates of σ provide a good basis for characterizing the across-subject average percentage of correct performance.

Fig. S1 shows the individual subjects' estimates of internal noise σ. The values are very similar across most subjects (all except for subjects S4 and S12), confirming that they could reliably discriminate distance in our simulated auditory environment. The relatively worse performance of subject S12 (and, to a lesser extent subject S4) could be due to the technical limitations of our simulations or due to other factors, e.g., subjects' inability to follow the instructions.

A behavioral measurement was also performed during the fMRI experiments. During the experiments, the subject's vigilance and focus needed to be constant across the different stimulus conditions to prevent vigilance-related changes in brain activation from confounding the stimulus-related changes. To establish that different stimulus conditions were approximately equally demanding, 50% of trials during imaging experiments contained deviants, noise bursts of duration that was half of the standard bursts (Fig. 1*C*). The listeners were instructed to respond with a button press whenever they detected the deviant. Analysis of hit rates and reaction times was performed. One subject (S12) did not perform the task correctly (hit rate 0–20%). For the remaining subjects, the task difficulty was similar across the three different stimulus types (across-subject average hit rate of 93.5, 93.2, and 91% and reaction times of 630, 660, and 660 ms for the varying distance, varying intensity, and constant stimuli, respectively), suggesting that any fMRI activation differences across conditions cannot be attributed to differences in task difficulty. Subject S12 failed to follow the instructions for the behavioral task during the imaging experiments and also had the worst performance in the behavioral experiment. Therefore, this subject's imaging data were excluded.

Acoustic Analysis of Stimuli. Our behavioral results suggested that subjects used a combination of available cues, instead of direct to reverberant ratio (D/R) or interaural level difference (ILD) alone, to discriminate sound-source distances. However, an additional acoustic analysis of the present stimuli was conducted to examine whether either one of the individual cues, considered separately, provided information consistent with the observed distance discrimination. This analysis was also necessary because, unlike D/R, ILD also contributes to direction discrimination, thus suggesting an alternative hypothesis that our main results reflect a partial byproduct of ILD-specific direction neurons, instead of populations encoding distance, per se.

This acoustic analysis of D/R and ILD is presented in Fig. S2. To generate the graphs, the BRIRs used in the simulations were first divided into their direct-sound part, corresponding to the sound that a listener would hear in an anechoic environment, and their reverberant part, corresponding to the reflections heard in the room. Then, the complete BRIRs as well as the separate impulse responses were convolved with random noise tokens corresponding to the stimuli. The noises were either broadband stimuli, as used in the experiments, or one-third octave-filtered narrowband stimuli, with the filters centered at 450 Hz or 6,500 Hz, marked in Fig. S2 as Broad, Low, and High, respectively. Root-mean-square energy was computed in the convolved stimuli and differences in this energy were plotted as D/Rs or ILDs. D/Rs were computed separately for the near ear (ipsilateral to the simulated target location) and far ear (contralateral to the target). ILDs were computed either for the total stimulus (including direct and reverberant part) or for the direct part only.

Fig. S2, *Left* shows that, as expected, the D/Rs tend to decrease as the stimulus moves away from the listener. The decrease is much steeper at the near ear than at the far ear (compare the thick solid and dotted lines). This difference in steepness confirms that the near-ear D/R is a more informative cue than the far-ear D/R. At the near ear, this decrease is approximately linear, independent of the stimulus frequency/bandwidth (compare the three red solid lines in the D/R panel). Furthermore, although the absolute D/R differs across frequencies, the slopes of three D/R graphs for the near ear are very similar to each other, changing linearly and spanning ∼20 dB as the stimulus moves from 15 to 100 cm. That is, a particular relative change in the sound-source distance (say, 50 vs. 100 cm) is represented by a consistent relative D/R change at all frequencies.

The right-hand panel in Fig. S2 shows the results for the ILDs. The pattern of results is more complex than for the D/Rs. First, as with D/Rs, the ILDs decrease as the stimulus moves away from the listener. However, unlike with D/R, this decrease is nonlinear as a function of stimulus distance (steeper for near than far sources) and it is steeper at high frequencies (spanning ~25 dB, symbols "Δ") than at low frequencies (spanning only about 15 dB, symbols "∇"). Similarly, the absolute values of the ILDs are larger at high frequencies than at low frequencies. Another important aspect is revealed by the comparison of the total and direct broadband ILDs (compare the solid and dotted lines). Whereas reverberation has a very small effect on ILD for very near sources, it results in a larger reduction of ILDs for distant sources.

In summary, these results show that, if distance perception were based on a single cue, the near-ear D/R would be a more parsimonious cue in reverberation than ILD. This is the case because there is a direct linear relationship between D/R and distance, and the slope of this linear relationship is similar across frequencies. Thus, one can always perform a relative distance judgment on the basis of D/R, independent of the stimulus frequency content or the distances considered. This result is not surprising, as the reverberant field is approximately constant for a stimulus presented at a fixed level, whereas the direct sound level is simply inversely proportional to the stimulus distance. For example, the D/R panel in Fig. S2 shows that a decrease in D/R by ~20 dB always corresponds to quadrupling of the stimulus distance. On the other hand, to perform a relative distance judgment on the basis of ILD, different ILD-to-distance relationships apply depending on stimulus frequency (ILD changes more at high frequencies), the reference distance (ILD changes more for nearby sources), and on the level of reverberation in a given room (the more reverberant the room, the more the ILD changes with distance). The dependence of ILDs on frequency, distance, and reverberation level is particularly important from the neurophysiological perspective, as it contradicts the scenario that at the final processing stages sound distances are estimated by a fixed set of neurons tuned specifically to ILD only. If ILD were the dominating cue, there would have to be a complex arrangement of frequency-specific, distance-specific, and reverberation-level–specific neurons tuned to a variety of ILDs. In other words, although the average ILD of a broadband sound (thick black line in Fig. S2) decreases monotonically, the lack of consistency across frequencies, distance, and reverberation level makes it unlikely that such a cue could be used, consistently and reliably, for distance discriminations, particularly because much better cues related to the D/R are concurrently available in natural listening situations. Finally, as also noted in the main results, the discrimination sensitivity for ILD, on the basis of Weber's law, would differ from the present observations.

Predictions of Distance Discrimination. The distance discrimination analysis of the behavioral experiment found that performance was independent of the absolute distance of sources as long as the relative distance was fixed (e.g., the listener's ability to discriminate 25- vs. 50-cm sources was the same as the ability to discriminate 50- vs. 100-cm sources; Fig. 2). That is, subjects' ability to discriminate simulated distances was predicted by Weber's law, a principle that has been well documented to apply to many aspects of human perceptual sensitivity. This would be consistent with an idea that the combination of available cues, constituting the building blocks of a sound-distance percept, was used by the subjects. The purpose of this supporting discussion and analysis is to discuss predictions of how subjects would have performed, as predicted by Weber's law, if discriminations had been based on either D/R or ILD alone.

Kopčo and Shinn-Cunningham (3) performed a computational acoustic analysis of a behavioral experiment in which absolute distance responses were reported and concluded that all of the results of that study could be explained by a model that assumed that the listeners only used the D/R cue for their judgments. For the current data this model would predict that the sensitivity in distance discrimination is dependent on only (*i*) the rate with which the D/R varies as a function of the source distance and (*ii*) the perceptual sensitivity (just noticeable difference, JND) to variations in D/R. The dependence of D/R on distance is linear (as shown in Fig. S2). The perceptual sensitivity to changes in D/R has only been studied by Larsen et al. (4) who found that it is fairly nonlinear for broadband stimuli as used here. Specifically, Larsen et al. measured D/R JNDs at baseline D/Rs of −10, 0, 10, and 20 dB, and found JNDs of ~3 dB at 0 and 10 dB and JNDs of 6–8 dB at −10 and 20 dB . Given that the rate of change in D/R with distance is constant and considerably larger than 10 dB, except for the most distant sources (Fig. S2), this pattern of JNDs suggests that the listeners should be more sensitive to changes in distance at the largest distances examined here (in particular when the low-frequency channels are considered). However, no such difference in distance discrimination performance was observed here.

A similar model can be used to predict discrimination performance if subjects only used ILDs. Fig. S2 shows that the rate of change in ILDs with distance is highest for nearby sources. Because ILD JNDs are approximately constant at 1 dB (5), the best distance discrimination performance based on ILDs would be expected at the nearest distances, which is not consistent with our main results (Fig. 2).

Taken together, these results suggest that, whereas it may still be the most parsimonious assumption that performance is based on D/Rs, the process by which the brain computes distance estimates is more complex than a simple D/R-to-distance mapping. For example, D/R information might be combined across multiple channels or ILD information might also be used. Future studies will have to be performed to describe the mechanism of distance processing in more detail.

Supporting fMRI Experiment. The monaural D/R at the ear closer to the source (near ear) has been suggested to be the main intensity-independent distance cue used for nearby sources in reverberation (3). Therefore, a supporting experiment was performed in which presentation intensity was normalized such that the overall energy received at the near ear was constant, independent of the source distance. It was expected that, after the normalization, the intensity-independent D/R cue would dominate distance processing whereas activations related to variations in intensity would be minimized. The supporting experiment was identical to the main fMRI experiment, except that only two types of stimuli were used: constant vs. a near-ear normalized version of the varying distance stimulus (Fig. 1*C*). When D/R and ILD were varied randomly and intensity at the near ear was held constant, we observed a more widespread activation of auditory cortices (red-yellow in Fig. S3*A*) than in the main experiment. The observed activations are consistent with our interpretation that, compared with the main experiment, near-ear intensity normalization does not eliminate all intensity variations (such as the far-ear overall intensity variation, near-ear/far-ear direct-sound intensity variation, and reverberant-sound intensity variations). More specifically, as predicted by our hypothesis, the strongest "extra" activations were observed in the right hemisphere, that is, contralateral to the far ear for which the overall intensity was not normalized (Fig. S3*A*, *Right*). Nevertheless, in the hemisphere contralateral to the stimuli, the voxel with strongest activation was located in planum temporale (PT), in the posterior nonprimary auditory cortex (surface Talairach $\{x, y, z\} = \{-50, -30, 8\}$), consistent with our main result and the predictions of the dual pathway model of auditory cortex (6–10) (Fig. S3*B*).

1. Zahorik P (2002) Assessing auditory distance perception using virtual acoustics. *J Acoust Soc Am* 111:1832–1846.
2. Durlach NI, Braida LD (1969) Intensity perception. I. Preliminary theory of intensity resolution. *J Acoust Soc Am* 46:372–383.
3. Kopčo N, Shinn-Cunningham BG (2011) Effect of stimulus spectrum on distance perception for nearby sources. *J Acoust Soc Am* 130:1530–1541.
4. Larsen E, Iyer N, Lansing CR, Feng AS (2008) On the minimum audible difference in direct-to-reverberant energy ratio. *J Acoust Soc Am* 124:450–461.
5. Hartmann WM, Constan ZA (2002) Interaural level differences and the level-meter model. *J Acoust Soc Am* 112:1037–1045.
6. Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268:111–114.
7. Rauschecker JP (1997) Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Otolaryngol Suppl* 532(Supplement):34–38.
8. Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci USA* 97:11800–11806.
9. Rauschecker JP (1998) Cortical processing of complex sounds. *Curr Opin Neurobiol* 8: 516–521.
10. Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. *Science* 292:290–293.

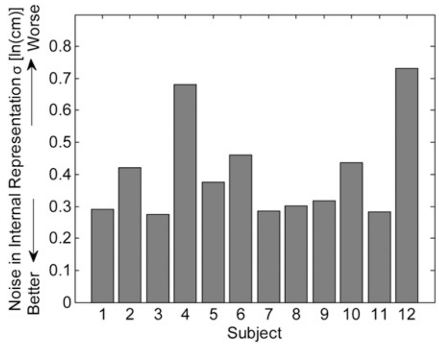

Fig. S1. Individual listeners’ sensitivity to distance changes. Each bar represents one subject’s estimate of the noise (i.e., variance) in his/her internal representation of sound-source distance on the basis of a decision-theory model fit to that subject’s individual percentage of correct data from the behavioral experiment. Performance of subjects S4 and S12 was considerably worse than for the remaining subjects. As an example, on the basis of these sensitivity estimates, subject S12 is expected to correctly discriminate the two most separated speakers only in less than 90% of trials, whereas the well-performing subjects would achieve accuracy of about 99%.

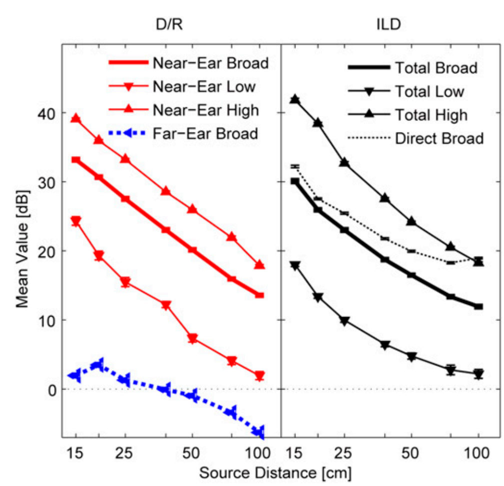

Fig. S2. Mean D/R (*Left*) and ILD (*Right*) as a function of distance for the stimuli used in this study. Cue values were computed either for the whole broadband stimulus (Broad), or in one-third octave bands centered at 450 Hz (Low) and 6,500 Hz (High). D/Rs are shown for the ear ipsilateral to the stimulus (Near Ear) or contralateral to the stimulus (Far Ear). ILDs were computed either for the whole stimulus (Total; including both the direct sound and reverberation) or for the direct sound only (Direct; corresponding to the ILD that would be obtained in an anechoic environment). Error bars represent SDs across random noise tokens used as stimulus.

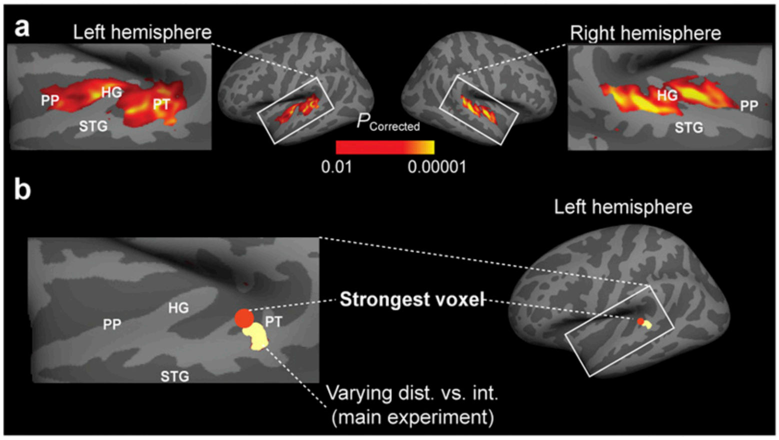

Fig. S3. (*A*) Adaptation-fMRI data from the follow-up fMRI experiment presented on inflated cortical surface representations. Contrast shows activation for sounds varying in distance with intensity normalized at the near ear vs. nonvarying standard sounds. (*B*) Location of the strongest-activated voxel in the hemisphere contralateral to the stimulation direction. The voxel is very near to the significant activation cluster from varying distance vs. varying intensity contrast of the main experiment, presumably revealing intensity-independent distance-cue representations in the posterior nonprimary auditory cortices.