

Human Exploration of Enclosed Spaces through Echolocation

Virginia L. Flanagan,^{1*}  Sven Schörnich,^{2*} Michael Schraner,² Nadine Hummel,¹ Ludwig Wallmeier,² Magnus Wahlberg,³  Thomas Stephan,^{1,4} and  Lutz Wiegrebe²

¹German Center for Vertigo and Balance Disorders, Ludwig-Maximilians-Universität (LMU Munich), Munich, Germany, ²Division of Neurobiology, Department of Biology II, LMU Munich, Munich, Germany, ³Fjord&Bælt and University of Southern Denmark, DK-5300 Kerteminde, Denmark, and ⁴Department of Neurology, LMU Munich, Munich, Germany

Some blind humans have developed echolocation, as a method of navigation in space. Echolocation is a truly active sense because subjects analyze echoes of dedicated, self-generated sounds to assess space around them. Using a special virtual space technique, we assess how humans perceive enclosed spaces through echolocation, thereby revealing the interplay between sensory and vocal-motor neural activity while humans perform this task. Sighted subjects were trained to detect small changes in virtual-room size analyzing real-time generated echoes of their vocalizations. Individual differences in performance were related to the type and number of vocalizations produced. We then asked subjects to estimate virtual-room size with either active or passive sounds while measuring their brain activity with fMRI. Subjects were better at estimating room size when actively vocalizing. This was reflected in the hemodynamic activity of vocal-motor cortices, even after individual motor and sensory components were removed. Activity in these areas also varied with perceived room size, although the vocal-motor output was unchanged. In addition, thalamic and auditory-midbrain activity was correlated with perceived room size; a likely result of top-down auditory pathways for human echolocation, comparable with those described in echolocating bats. Our data provide evidence that human echolocation is supported by active sensing, both behaviorally and in terms of brain activity. The neural sensory-motor coupling complements the fundamental acoustic motor-sensory coupling via the environment in echolocation.

Key words: echolocation; fMRI; sensory-motor coupling; spatial processing; virtual acoustic space

Significance Statement

Passive listening is the predominant method for examining brain activity during echolocation, the auditory analysis of self-generated sounds. We show that sighted humans perform better when they actively vocalize than during passive listening. Correspondingly, vocal motor and cerebellar activity is greater during active echolocation than vocalization alone. Motor and subcortical auditory brain activity covaries with the auditory percept, although motor output is unchanged. Our results reveal behaviorally relevant neural sensory-motor coupling during echolocation.

Introduction

In the absence of vision, the only source of information for the perception of far space in humans comes from audition. Com-

plementary to the auditory analysis of external sound sources, blind individuals can detect, localize, and discriminate silent objects using the reflections of self-generated sounds (Rice, 1967; Griffin, 1974; Stoffregen and Pittenger, 1995). The sounds are produced either mechanically (e.g., via tapping of a cane) (Burton, 2000) or vocally using tongue clicks (Rojas et al., 2009). This type of sonar departs from classical spatial hearing in that the listener is also the sound source (i.e., he or she must use his or her own motor commands to ensonify the environment). It is a specialized form of spatial hearing also called echolocation that is known from bats and toothed whales. In these echolocating species, a correct interpretation of echo information involves precise sensory-motor coupling between vocalization and audition (Schuller et al., 1997; Smotherman, 2007). However, the importance of sensory-motor coupling in human echolocation is unknown.

Received March 30, 2012; revised Nov. 23, 2016; accepted Dec. 1, 2016.

Author contributions: V.L.F., S.S., M.S., T.S., and L. Wiegrebe designed research; V.L.F., S.S., M.S., N.H., L. Wallmeier, T.S., and L. Wiegrebe performed research; V.L.F., S.S., M.S., N.H., L. Wallmeier, T.S., and L. Wiegrebe analyzed data; V.L.F., S.S., M.S., M.W., T.S., and L. Wiegrebe wrote the paper.

This work was supported by the Deutsche Forschungsgemeinschaft Wi 1518/9 to L. Wiegrebe and German Center for Vertigo and Balance Disorders BMBF IFB 01EO0901 to V.L.F. We thank Daniel Kish for insightful discussions; and the World Access for the Blind and Benedikt Grothe for providing exceptionally good research infrastructure and thoughtful discussions on the topic.

The authors declare no competing financial interests.

*V.L.F. and S.S. contributed equally to this study.

Correspondence should be addressed to Dr. Virginia L. Flanagan, Neurological Research Pavilion, University Hospital Munich Großhadern, Feodor-Lynen-Str. 19, 81377 Munich, Germany. E-mail: vflanagin@lrz.uni-muenchen.de.

DOI:10.1523/JNEUROSCI.1566-12.2016

Copyright © 2017 the authors 0270-6474/17/371614-14\$15.00/0

Neuroimaging studies on echolocation have shown that the presentation of spatialized echoes to blind echolocation experts results in strong activations of visual cortical areas (Thaler et al., 2011, 2014b). In these studies, participants did not vocalize during imaging, an approach we will refer to as “passive echolocation.” While these studies have resulted in valuable insights into the representations in, and possible reorganizations of, sensory cortices, passive echolocation is not suitable to investigate the sensory-motor coupling of echolocation.

Sonar object localization may involve the processing of interaural time and level differences of echoes, similar to classical spatial hearing. For other echolocation tasks, however, the relative difference between the emitted vocalization and the returning echoes provides the essential information about the environment (Kolarik et al., 2014). Sonar object detection is easier in a room with reflective surfaces (Schenkman and Nilsson, 2010), suggesting that reverberant information, such as an echo, provides important and ecologically relevant information for human audition. Reverberant information can be used to evaluate enclosed spaces in passive listening (Seraphim, 1958, 1961), although it is actively suppressed when listening and interpreting speech or music (Blauert, 1997; Litovsky et al., 1999; Watkins, 2005; Watkins and Makin, 2007; Nielsen and Dau, 2010). Psychophysical analyses of room size discrimination based on auditory information alone are scarce (McGrath et al., 1999).

How can we quantify the acoustic properties of an enclosed space? The binaural room impulse response (BRIR), a measure from architectural acoustics of the reverberant properties of enclosed spaces (Blauert and Lindemann, 1986; Hidaka and Beranek, 2000), captures the complete spatial and temporal distribution of reflections that a sound undergoes from a specific source to a binaural receiver. In a recent study, we introduced a technique allowing subjects to actively produce tongue clicks in the MRI and evaluate real-time generated echoes from a virtual reflector (Wallmeier et al., 2015). Here, we used this same technique but with a virtual echo-acoustic space, defined by its BRIR, to examine the brain regions recruited during human echolocation.

In this study, participants vocally excited the virtual space and evaluated the echoes, generated in real-time, in terms of their spatial characteristics. They had full control over timing and frequency content of their vocalizations and could optimize these parameters for the given echolocation task. As such, we consider this active echolocation. First, we quantified room size discrimination behavior and its relationship to the vocalizations' acoustic characteristics. Then we compared the brain activity and performance between active and passive echolocation to elucidate the importance of active perception. The relationship between brain activity and the behavioral output was investigated in a parametric analysis. Finally, we compared the brain activity of a blind echolocation expert during active echolocation to the sighted subjects we measured.

Materials and Methods

Three experiments on active echolocation in humans were performed. First, a psychophysical experiment (see Room size discrimination) examined the effect of individual call choice on performance. Second, we examined the difference between active and passive echolocation in terms of behavior and brain activity, as measured with fMRI (see Active vs passive echolocation). Finally, we tested the relationship between brain activity and perceived room size in a group of sighted subjects and in a blind echolocation expert (see Active echolocation only and Blind echolocation expert). The acoustic recordings and stimuli were the same for all three experiments and will be explained first. All experiments were

approved by the ethics committee of the medical faculty of the LMU (Projects 359-07 and 109-10). All participants gave their informed consent in accordance with the Declaration of Helsinki and voluntarily participated in the experiment.

Acoustic recordings. To conduct the experiments under realistic conditions, the BRIR of a real building was measured. A small chapel in Gräfelfing, Germany (Old St. Stephanus, Fig. 1A) with highly reflective surfaces was chosen because the reverberation time (i.e., the time it takes for reflections of a direct sound to decay by 60 dB) was long enough to not be masked by the direct sound. The floor consisted of stone flaggings; the walls and the ceiling were made of stone with wall plaster and the sparse furnishings were wooden. The chapel had a maximum width of 7.18 m, a maximum length of 17.15 m and a maximum height of 5.54 m.

BRIR recordings were performed with a B&K head-and-Torso Simulator 4128C (Brüel & Kjaer Instruments) positioned in the middle of the chapel facing the altar (Fig. 1A). Microphones of the head-and-torso simulator were amplified with a Brüel & Kjaer Nexus conditioning amplifier. The recording was controlled via a notebook connected to an external soundcard (Motu Traveler). The chapel was acoustically excited with a 20 s sine sweep from 200 to 20,000 Hz. The sweep was created with MATLAB (The MathWorks); playback and recording were implemented with SoundMxPro (HörTech). The frequency response of the mouth simulator was digitally equalized. The sweep was amplified (Stereo Amplifier A-109, Pioneer Electronics) and transmitted to the inbuilt loudspeaker behind the mouth opening of the head-and-torso simulator. The BRIR was extracted through cross-correlation of the emission and binaural recording (Fig. 1B) and had a reverberation time of ~ 1.8 s. This BRIR recording was used for all of the following experiments.

Stimuli. The BRIRs presented were all derived from the BRIR recorded in the chapel (see Acoustic recordings). The BRIRs were compressed along the time axis, a technique well established for scale models in architectural acoustics (Blauert and Xiang, 1993), resulting in scaled-down versions of the original, measured space. The BRIR recorded in the chapel was compressed by factors 0.2, 0.5, and 0.7; a compression factor of 0.2 produced the smallest room. The reverberation time scales with the same compression factors. From these reverberation times, the volume of a cube that would produce an equal reverberation time can be calculated according to Sabine (1923) (compare Fig. 1C). The spectral center of gravity of the BRIR increases with decreasing compression factor (Fig. 1C). The covariation of spectral and temporal parameters of the BRIRs is characteristic of the reverberations from different-sized rooms. Also, the overall level of the BRIR decreases with temporal compression: specifically, attenuations were -2 , -3 , and -9 dB for compression factors of 0.7, 0.5, and 0.2, respectively.

The experimental setup was designed around a real-time convolution kernel (Soundmexpro) running on a personal computer (PC with Windows XP) under MATLAB. Participants' vocalizations were recorded, convolved with a BRIR, and presented over headphones in real time, with the echo-acoustically correct latencies.

The direct sound (i.e., the sound path from the mouth directly to the ears) was simulated as a switchable direct input-output connection with programmable gain ('asio direct monitoring') with an acoustic delay of <1 ms. The result of the real-time convolution was added with a delay equal to the first reflection at 9.1 ms. The correct reproduction of the chapel acoustics was verified using the same recording setup and procedure as in the chapel but now the head-and-torso simulator was equipped with the experimental headset microphone and earphones in an anechoic chamber (see Psychophysical procedure).

Room size discrimination

Here, we psychophysically quantified the ability of sighted human subjects to detect changes in the size of an enclosed space by listening to echoes of their own vocalizations.

Participants. Eleven healthy subjects with no history of medical or neurological disorder participated in the psychophysical experiment (mean \pm SD age 23.4 ± 2.2 years; 4 female).

Procedure. The psychophysical experiments were conducted in a $1.2 \text{ m} \times 1.2 \text{ m} \times 2.2 \text{ m}$ sound-attenuated anechoic chamber

(G+H Schallschutz). Just noticeable differences (JND) in acoustic room size were quantified using an adaptive two-interval, two-alternative, forced-choice paradigm. Each observation interval started with a short tone beep (50 ms, 1000 Hz) followed by a 5 s interval in which both the direct path and the BRIR were switched on. Within this interval, subjects evaluated the virtual echo-acoustic space by emitting calls and listening to the echoes from the virtual space. The calls were typically tongue clicks (see Results; Fig. 3). The end of an interval was marked by another tone beep (50 ms, 2000 Hz). The pause between the two intervals of each trial was 1 s. After the end of the second interval, the subjects judged which of the two intervals contained the smaller virtual room (smaller compression factor). To focus the subjects' attention away from overall loudness toward the temporal properties of the reverberation, we roved the amplitude of the BRIR by ± 6 dB across intervals. This rove rendered discrimination based on the sound level of the reverberation difficult, at least for the larger three compression factors (see Stimuli, above).

Subjects were equipped with a professional headset microphone (Sennheiser HS2-EW) and in-ear headphones (Ety-motic Research ER-4S). The headset microphone was positioned at a distance of ~ 3 cm to the left of the subjects' mouth. Headphones and microphone were connected to an external soundcard (RME Fireface 400), which was connected to the PC. A gamepad (BigBen interactive) was used as response device. Auditory feedback was provided with a 250 ms tonal sweep, which was upward modulated for a correct decision and downward modulated for a wrong decision.

Compression-factor JNDs were measured following a three-down, one-up rule (i.e., the difference between the two intervals was reduced after three correct decisions and increased after one incorrect decision). An adaptive track was continued until 11 reversals (a wrong response after three consecutive correct trials, or three correct responses after one wrong response) were gathered. The compression-factor difference was 2 for reversals 1–3, 1.2 for reversals 4 and 5, and 1.1 for reversals 6–11. The mean compression-factor difference across the last six reversals was taken as the threshold for an experimental run. Data shown are the average of three consecutive runs, once the subjects' performance was stable (i.e., the SD of the thresholds across the last three runs was less than $\frac{1}{4}$ of the mean threshold). JNDs are specified by the percentage of each side of

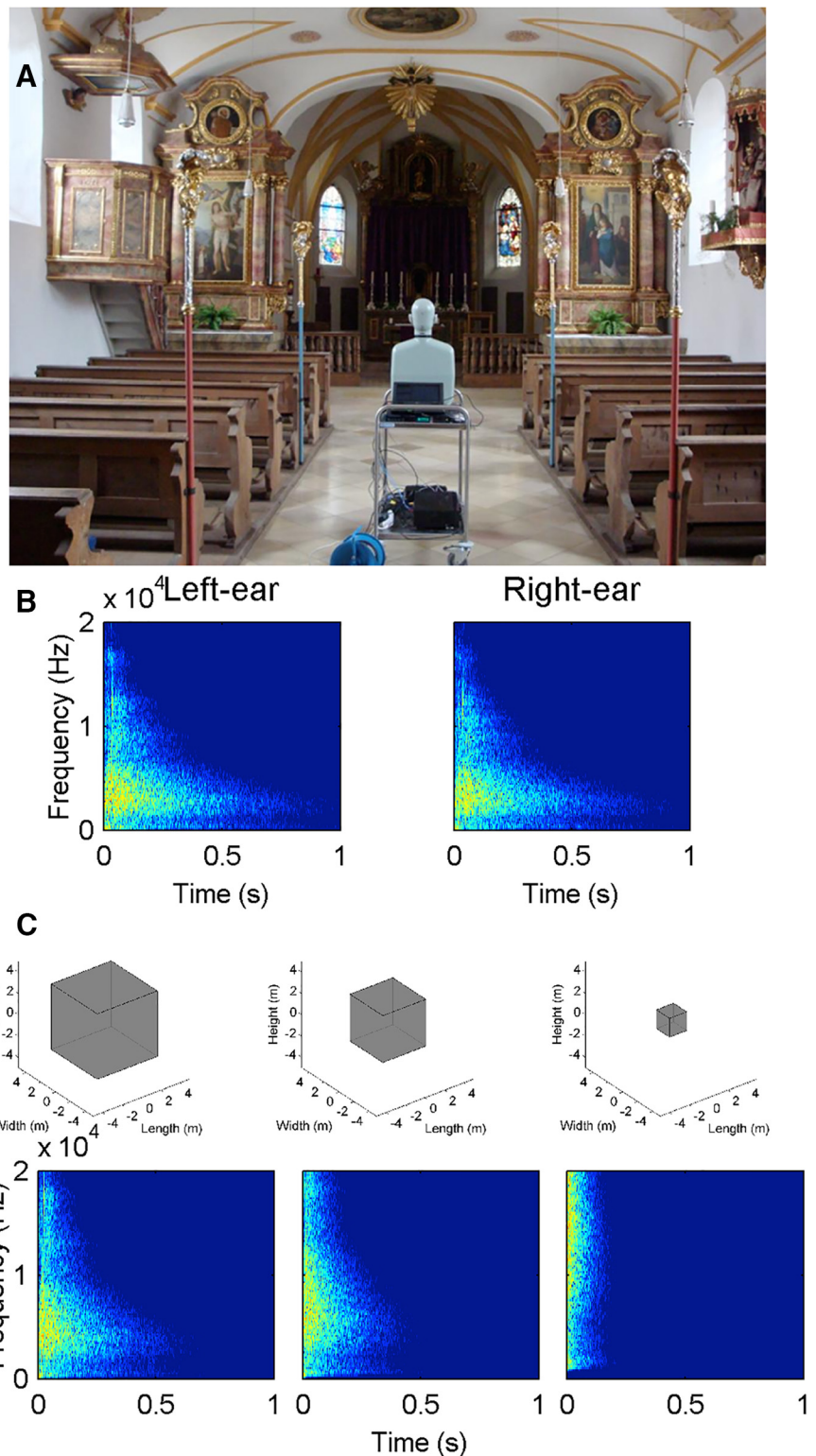


Figure 1. BRIR of a real enclosed space. **A**, Photograph of the acoustically excited room (Old St. Stephanus, Gräfelting, Germany) with the head-and-torso simulator. **B**, Spectrograms of the left and right BRIRs. Sound pressure level is color coded between -60 and 0 dB. **C**, Changes of the size of a virtual room with an equivalent reverberation time after it is compressed with factors of 0.7, 0.5, and 0.2, respectively. Bottom, Spectrograms of the left-ear room impulse response corresponding to the three compression factors. Color scale is identical to the second row.

the virtual room that must be increased such that the BRIR changes perceptibly.

The psychophysical procedure challenged the subjects to optimize both their vocal emissions and the auditory analysis of the

virtual echoes to extract room size-dependent echo characteristics based on the trial-to-trial feedback. Considering that loudness, spectral, and temporal cues covaried with IR compression, we cannot isolate the perceptual cue or combination of cues that was used. However, listeners were deterred from using loudness cues by the roving-level procedure. Parts of the current psychophysical data were presented at the 2012 International Symposium on Hearing and can be found in the corresponding proceedings (Schörmich et al., 2013).

Sound analysis. To test for the effects of individual sound vocalizations on psychophysical performance, we analyzed the temporal and spectral properties of the echolocation calls used by each subject. The microphone recording from the second interval of every fifth trial was saved to hard disk for a total number of available recordings per subject of between 300 and 358. The number of calls, RMS sound level, duration, and frequency were analyzed from these sound recordings. The number of calls in each recording was determined by counting the number of maxima in the recording's Hilbert envelope that exceeded threshold (mean amplitude of the whole recording plus three times the SD of the amplitude). Clipped calls and calls starting within the last 50 ms of the interval were excluded from further analysis. Each identified call was positioned in a 186 ms rectangular temporal window to determine the RMS sound level. The call duration was determined as the duration containing 90% of the call energy. The peak frequency of each call was determined from the Fourier transform of the 186 ms rectangular window. Correlations between an echolocation-call parameter of a subject and that subjects' JND were quantified using Spearman's ρ .

Active versus passive echolocation

To understand the importance of active sensing for echolocation, we compared active and passive echolocation while measuring brain activity with fMRI. In this experiment, participants judged the size of a virtual room by either actively producing vocalizations or passively listening to previously produced vocalizations and evaluating the resulting echoes.

Participants. Ten healthy participants with no history of medical or neurological disorder took part in the experiment (age 25.2 ± 3.1 years; 6 females). Three subjects from the room size discrimination experiment participated in this experiment. All participants were recruited from other behavioral echolocation experiments to ensure that they were highly trained in echolocation at the time of the experiment.

Setup. During active echolocation, subjects preferred calls were recorded by an MRI-compatible optical microphone (Sennheiser MO 2000), amplified (Sennheiser MO 2000), converted (Motu Traveler), convolved in real time with one of four BRIRs, converted back to analog (Motu Traveler), and played back over MRI compatible circumaural headphones (Nordic Neurolabs). The frequency-response characteristics of this setup were calibrated with the head-and-torso simulator to ensure that the BRIR recorded with the MRI-compatible equipment was identical to the BRIR measured with the same simulator in the real (church) room. The convolution kernel and programming environment were the same as for the psychophysics experiment.

Procedure. The task was to rate the size of the room on a scale from 1 to 10 (magnitude estimation) when presented with one of four BRIR compression factors (see Stimuli). Subjects were instructed to close their eyes, to keep their heads still, and to use a constant number of calls for each trial. A single trial consisted of a 5 s observation interval, where subjects produce calls and evaluate the virtual echoes, bordered by auditory cues (beeps to de-

lineate the start and end of an observation interval). Passive and active trials were signaled to the subjects with beeps centered at 0.5 and 1 kHz, respectively. The observation interval was temporally jittered within a 10 s window across repetitions (0.4–4.8 s from the start of the window in 0.4 steps). The 10 s window allowed us to provide a quiescent period for the task, followed by one MRI acquisition. Jittering was done to improve the fit of the functional imaging data by sampling from different points of the hemodynamic response function and is a way to optimize sampling of the hemodynamic signal.

The time from the start of the 5 s echolocation interval to the start of fMRI acquisition was therefore between 10.1 and 5.7 s. Following the 10 s window, one MR image (2.5 s) was collected framed by two 500 ms breaks after which subjects verbally expressed their rating within a 3 s response interval bordered by 2 kHz tone beeps. The total trial time was 16.5 s.

In half of the trials, participants actively vocalized (active echolocation) and in half of the trials calls and echoes were passively presented to the participants (passive echolocation). In the passive trials, vocalizations of a randomly chosen, previously recorded active trial was convolved with a BRIR, and presented to participants. Thus, in the passive trials, subjects received the same auditory input as in a previous active trial, but the subject did not vocalize. Three additional null-conditions were introduced: (1) an active-null during which subjects vocalized, but neither direct sound nor echoes were played through the headphones; (2) a passive-null in which the previously recorded vocalizations were presented through an anechoic BRIR; and (3) silence (complete-null), in which no sound was presented and no vocalizations were made. This resulted in a total of 5 active conditions (four BRIRs and one null), 5 passive conditions (four BRIRs and one null), and a complete-null condition. All null conditions were to be rated with a "0."

In a 40 min session, subjects were trained on the timing of the procedure and to distinguish between active and passive trials. One MRI session included two runs of fMRI data acquisition. Within one run, the 11 pseudo-randomized conditions were repeated five times, for a total of 55 trials in each run. Subjects were scanned in two separate sessions for a total of four runs of fMRI data acquisition.

Image acquisition. Images were acquired with a 3T MRI Scanner (Signa HDx, GE Healthcare) using a standard 8-channel head coil. The 38 contiguous transverse slices (slice thickness 3.5 mm, no gap) were acquired using a gradient echo EPI sequence (TR 16.5 s, TE 40 ms, flip angle 90 deg, matrix 64×64 voxel, FOV 220 mm, interleaved slice acquisition). Image acquisition time was 2.5 s; the remaining 14 s of quiescence minimized acoustical interference during task performance, a methodological procedure known as sparse imaging (Hall et al., 1999; Amaro et al., 2002). A T1-weighted high-resolution structural image of the entire brain ($0.8 \times 0.8 \times 0.8$ isotropic voxel size) was also acquired using a fast spoiled gradient recalled sequence.

Analysis. To test for behavioral performance differences between active and passive echolocation, a within-subject 2×4 ANOVA with factors active/passive and BRIR compression factor was performed. Two separate within-subject one-way ANOVAs were then used to assess whether loudness and number of clicks differed between BRIR compression factors.

Image processing and data analysis were performed using SPM8 (Wellcome Trust Centre for Neuroimaging, UCL, London) for MATLAB. Volumes were corrected for head motion using realignment, and spatially normalized to MNI space through segmentation of the high-resolution MR image (Ash-

burner and Friston, 2005). Images were smoothed with an 8 mm FWHM isotropic Gaussian kernel to reduce spatial noise.

Single-subject effects were tested with the GLM. High-pass filtering (cutoff time constant = 500 s) the time series reduced baseline shifts. Each run was modeled separately in one design to correct for within-run effects. The 5 s observation interval for active and passive echolocation trials and their null conditions (active-null, passive-null) were modeled separately as boxcar functions convolved with the hemodynamic response function. The four BRIRs were combined into a single regressor for either active or passive echolocation. In addition, two regressors corresponding to the mean centered linear parametric modulation of reported room size for active and passive trials separately, modeled additional variability in the experimental design. The complete silence null was not explicitly modeled. Head movement parameters were included as regressors of no interest.

The behavioral results of the room size rating task showed us that participants could not distinguish between the smallest BRIR compression factor (0.2) and the passive null, without echoes (see Results). Therefore, in the analyses, we did not use the passive null but compared passive echolocation to the baseline control null. The two contrast images corresponding to the subtractive effects of echolocation compared with null (active echolocation, active null; and passive echolocation, baseline) were used to create a paired *t* test at the group level to compare active and passive echolocation. Voxels exceeding an extent threshold of five contiguous voxels and a voxel-level height threshold of $p < 0.05$ corrected for multiple comparisons (false discovery rate [FDR]) (Genovese et al., 2002) were considered significant unless otherwise stated.

Active echolocation only

The active versus passive experiment randomly switched between active production of echolocation calls and passive listening to these calls. This task switching could have led to additional brain activation patterns that are not directly related to active or passive echolocation. We therefore performed a second experiment, in which participants only performed active echolocation during fMRI data acquisition. Throughout this experiment, all subjects actively produced consistent echolocation calls and were familiar with the vocal excitation and auditory evaluation of BRIRs.

Participants. The same participants that participated in the psychophysical experiment (see Room size discrimination) were recruited for this experiment.

Setup, imaging parameters, and procedure. The setup and the imaging parameters were the same as in the active versus passive echolocation experiment, except that only active echolocation trials were presented. A single trial consisted of a 5 s observation interval, where subjects produce calls and evaluate the virtual echoes, bordered by 2 kHz tone beeps. The observation interval was also temporally jittered within a 10 s window across repetitions (see Active versus passive echolocation). Each BRIR compression factor was additionally presented at four amplitude levels corresponding to 1, 2, 3, and 4 dB relative to the calibrated level. These small-level changes filled the level steps concomitant with the IR compression. The active-null condition, during which neither direct sound nor echoes were played through the headphones, was presented four times for every other combination of BRIR compression factor and amplitude level (16 combinations in total).

One scanning session included 2 runs of 3 repetitions, each repetition consisting of 20 pseudo-randomized trials (the 16 different reverberation conditions plus the four null conditions) for

a total of 60 trials per run, 48 of which were reverberation conditions. Subjects completed four runs in two separate sessions (each session was ~45 min).

Analysis. Room size ratings were analyzed using a within-subject 4×4 ANOVA with factors BRIR compression factor and amplitude level.

fMRI analysis, including preprocessing and significance levels, was the same as in the active versus passive echolocation experiment. For the single-subject GLMs, a single regressor was used to model all of the 16 conditions with echoes. The null condition was not explicitly modeled. Four additional regressors modeled linear and quadratic parametric modulations of the mean-centered room size rating and BRIR amplitude levels on each trial. Head movement parameters were included as regressors of no interest. Contrasts for echolocation-baseline and for the linear and quadratic modulations with BRIR amplitude levels and room size rating were entered into *t* tests at the group level.

No voxels were significantly correlated with the quadratic modulations of room size or BRIR amplitude. We therefore only report the linear modulations. We first compared the brain activity during active echolocation in the first experiment (active vs passive echolocation) to the activity during active echolocation in this experiment using a two-sample *t* test at the group level. We then tested the activation pattern during active echolocation compared with active vocalization without auditory feedback (the active null) using a one-sample *t* test.

Parametric modulations of brain activity with a stimulus or behavioral parameters (i.e., correlations between the strength of a stimulus or the response subjects and height of the brain activity) provide strong evidence that brain regions with significant parametric modulation are involved in the given task. Therefore, complementary to the subtractive analysis, we examined parametric modulations of brain activity with room size rating and BRIR amplitude changes using one-sample group-level *t* tests.

Blind echolocation expert. We also measured brain activity during echolocation of an echolocation expert to examine the brain regions recruited during active echolocation when audition is the primary source of information about far, or extrapersonal, space. The male congenitally blind, right-handed subject (age 44 years), performed the active only echolocation experiment, with the same imaging parameters, and single-subject data analysis. Additionally, a two-group model tested for significant differences in echolocation-null between the echolocation expert and the healthy subjects.

Results

Room size discrimination

All sighted subjects quickly learned to produce tongue clicks and perceive virtual rooms using echolocation. Subjects could detect changes in the BRIR compression factor independent of the roving BRIR amplitude levels, suggesting that they were able to use properties of the echo other than loudness to solve the task. The JNDs were quite stable within each subject but varied between about 5% and 25% across subjects. Previous findings on spatial acuity and object localization using echolocation in sighted subjects also found a high degree of variability in subjects' performance (Teng and Whitney, 2011). The across-subject mean is on the order of 10% (i.e., the percentage that each side of the virtual room must be increased to perceive a different sized room) (Fig. 2, bottom). To show what 10% means, we created theoretical rooms. The mean psychophysical performance was such that the gray-filled room could be discriminated from the transparent room surrounding it (compare Fig. 2, top). These discrimination

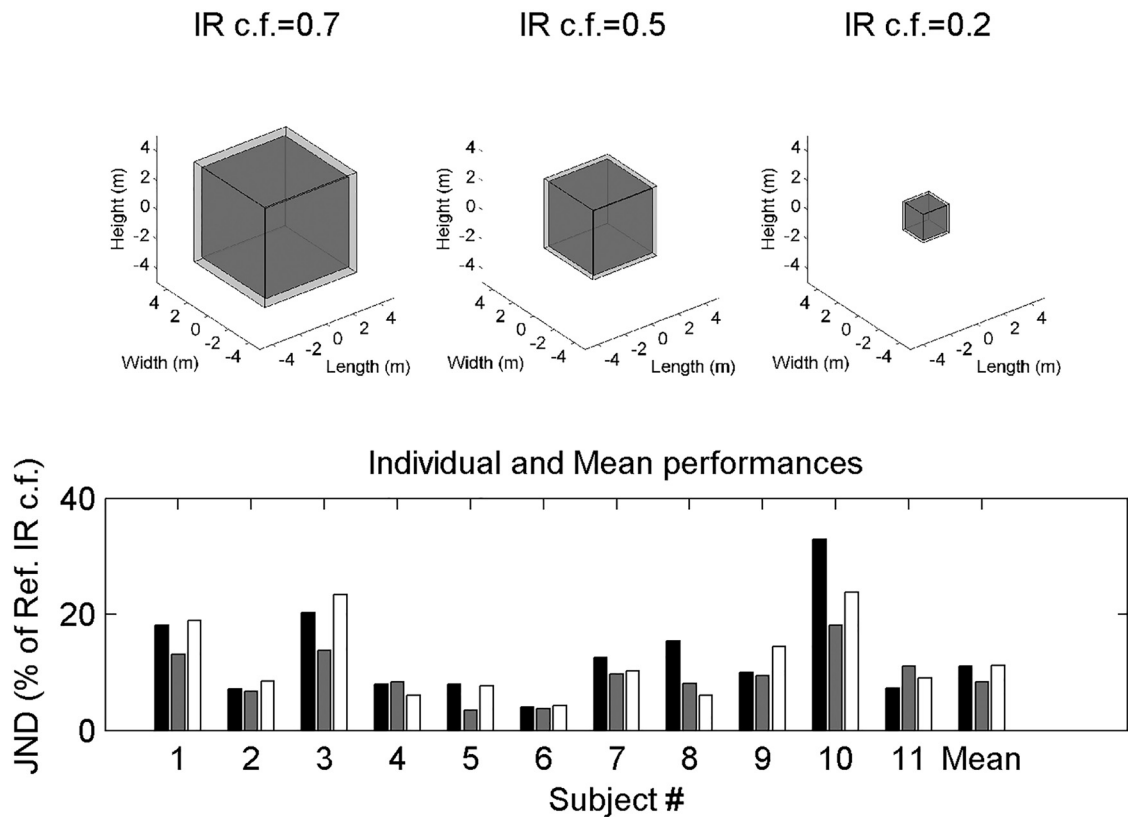


Figure 2. JNDs in room size. Top, Average JNDs illustrated in terms of the changes in the size of a cubic room with equivalent reverberation time (Sabine, 1923). Bottom, Individual JNDs are plotted for each subject and each room size and the mean on the far right. The individual data reveal that subjects performed quite differently, with some subjects having JNDs as low as 3%–4% and others having JNDs between 20% and 35%. Right, Across-subject mean. IR c.f., Impulse response compression factor.

thresholds were much finer than reported previously (McGrath et al., 1999) but are consistent with passive-acoustic evaluation of reverberation times (Seraphim, 1958).

Temporal and spectral call analyses revealed that all subjects produced relatively short, broadband tongue clicks at relative high sound levels to solve the psychophysical task (Fig. 3). Our participants, although free to choose their preferred vocalization, all produced clicks with durations that varied between 3 and 37 ms and absolute sound pressure levels (SPL) that varied between 88 and 108 dB SPL. The peak frequencies of the clicks ranged from 1 to 5 kHz. We then correlated the properties of the tongue-click with the JND for each subject to see which vocal-motor properties may be related to the psychophysical performance. Significant correlations were found between the click level and JNDs and the number of clicks per trial and JNDs, but there were no significant correlations for click duration and the peak frequency (Fig. 3, bottom). These effects do not survive a correction for multiple comparisons (for four independent tests); however, as the trends are in the same direction across all room sizes, this is likely due to the relatively small number of participants. Recruiting was an issue because of the time investment in training sighted subjects. Our results are also supported by previous work on the relationship between acoustic features of echolocation vocalizations and performance for object detection (Thaler and Castillo-Serrano, 2016).

In particular, in our study, louder clicks were associated with better JNDs than fainter clicks, presumably because the majority of the power from the echo is still above hearing thresholds (i.e., the virtual room is excited more effectively). A higher number of clicks per trial, on the other hand, corresponded to worse JNDs.

At first glance, this goes against the principle of “information surplus” (Schenkman and Nilsson, 2010); however, this effect is likely related to masking of the current reverberation by the subsequent click. Using short clicks or pulses with intermittent periods of silence, adjusted to target range, is also common in echolocating bats and toothed whales, allowing them to produce loud calls that effectively excite space and still analyze the comparatively faint echoes (Thomas et al., 2004). Humans trained to echolocate appear to optimize their vocalizations in a similar way.

Active versus passive echolocation

After characterizing performance psychophysically, we were interested in the brain activity during echolocation. Most of what we know about the neural basis of human echolocation is based on passive listening. Therefore, we first compared brain activation patterns between active-acoustic conditions, where subjects produced clicks in the scanner to passive-acoustic conditions where subjects only listened to clicks and their echoes. Data were collected with intermittent passive- and active-acoustic trials. Participants were asked to rate, on a scale from 1 to 10, the size of the virtual room, represented as one of four BRIR compression factors.

Behavioral performance

Both in the active and the passive-acoustic condition, subjects reliably rated the larger compression factors to correspond to a larger perceived room size (repeated-measures ANOVA, $F_{(4,36)} = 102.24, p = 7.44 \times 10^{-15}$; Fig. 4A). Although there was no main effect of echolocation type (active or passive) ($F_{(1,9)} = 0.015, p =$

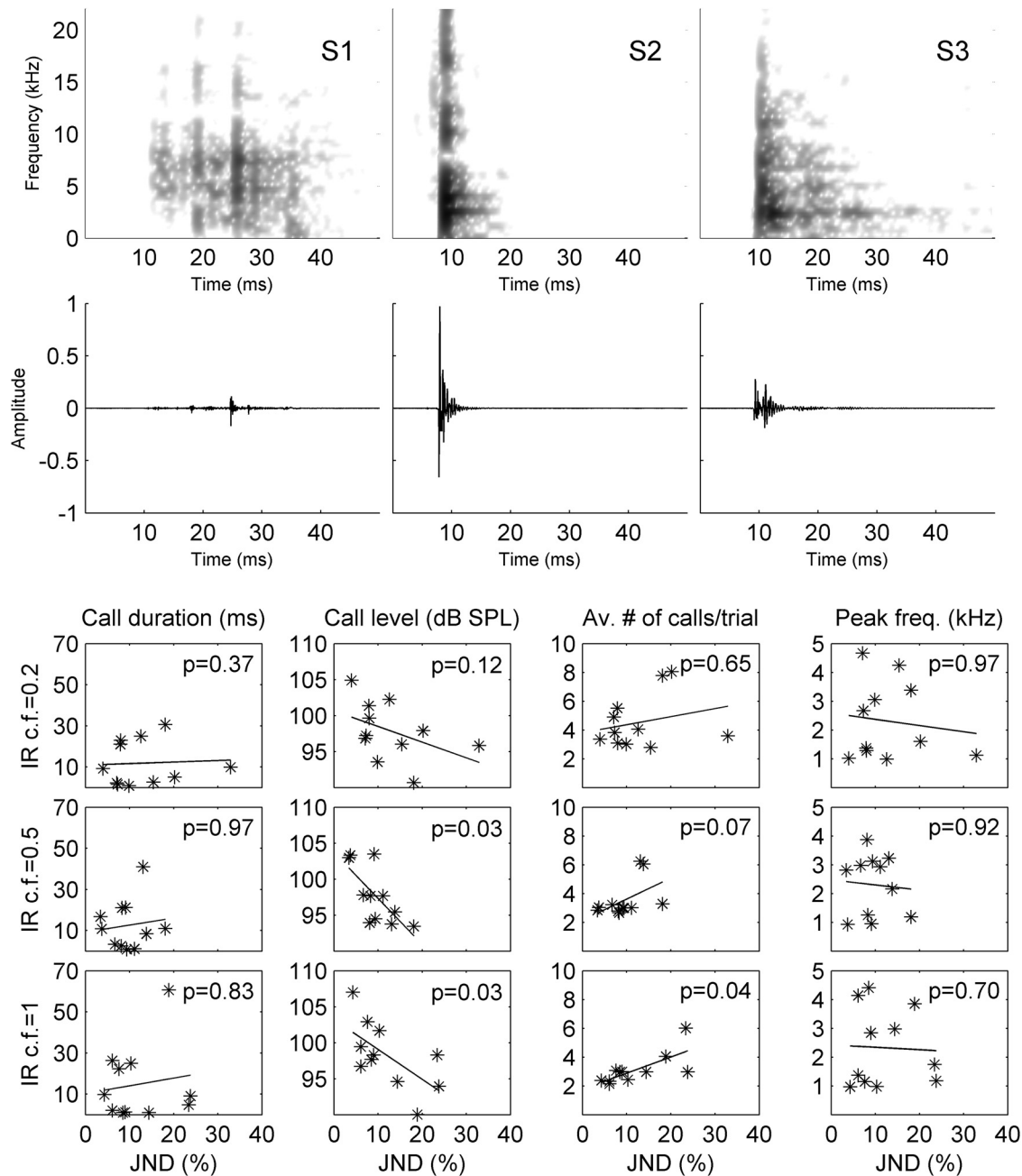


Figure 3. Examples of the subjects' vocalizations produced to solve the echo-acoustic task. Top, Exemplary spectrograms (Row 1) and oscillograms (Row 2) are shown for three typical participants. Bottom, A detailed correlation analysis between the individual psychophysical performances and specific call parameters. Each star corresponds to one subject. Top right, Correlation coefficients (Spearman's ρ). The analysis shows that, overall, JNDs improve with increasing call level and decrease with increasing number of calls per trial.

0.91), there was a significant interaction between room size and echolocation type ($F_{(4,36)} = 19.93, p = 5.11 \times 10^{-7}$). The ratings differed significantly across all active-acoustically presented compression factors but not across all passive-acoustically presented compression factors (Scheffé Test), and for the largest room, the active rating was significantly higher than the passive rating.

Subjects' vocalizations during the active-acoustic condition were also analyzed. Subjects produced between 9 and 10 clicks within each 5 s observation interval. The loudness and the number of clicks per observation interval did not differ significantly across the different compression factors or the null condition (ANOVA, $F_{(4,36)} = 0.41, p = 0.74$, and $F_{(4,36)} = 1.92, p = 0.15$,

respectively), confirming that subjects followed the instructions and did not attempt to change their motor strategy to aid in determining the room size.

Brain activity during active versus passive sensing

Because the number and loudness of clicks did not differ between the echolocation and the active null condition, any differences in brain activity between these two conditions in the motor cortices should be related to the sensory perception of the echoes from the BRIR compression factors and not the motor commands. To test for differences between active and passive echolocation, we compared active echolocation with the active null subtracted out, to passive echolocation. Signifi-

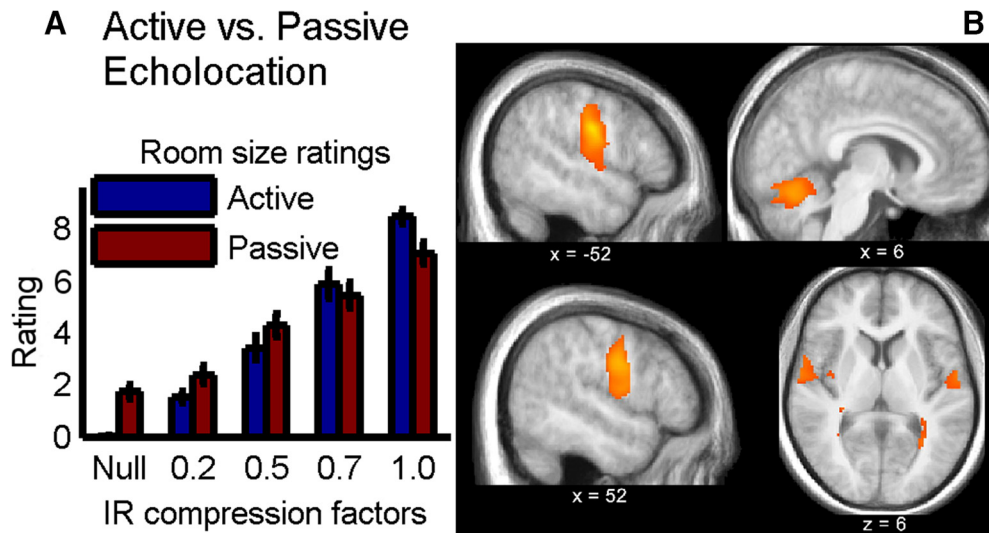


Figure 4. Active versus passive echolocation. **A**, Behavior: subjects’ rating of the perceived room size, in both active (blue) and passive (red) echolocation, for the four BRIR compression factors (room sizes). Error bars indicate SE across subjects. **B**, Neuroimaging: differential activations between active and passive echolocation show stronger motor activity during active echolocation, although the motor behavior was subtracted from the activity: (active echolocation – active null condition) – (passive echolocation – silence). Significant voxels ($p < 0.05$, FDR corrected) are shown as a heat map overlaid on the mean structural image from all subjects from the control experiment. Coordinates are given in MNI space (for details, see Table 1).

Table 1. Spatial coordinates of the local hemodynamic activity maxima for active echolocation: the active null, without auditory feedback, versus passive echolocation compared with the baseline null condition^a

Region	<i>x, y, z</i> (mm)	Z score	Extent
Cerebellum vermis V*	16, -64, -20	7.14	4634
	-20, -64, -20	6.69	
Postcentral gyrus, somatosensory cortex*	-56, -12, 26	6.83	4769
Precentral gyrus*	60, 2, 28	6.44	4159
Precuneus	-4, -40, -50	4.74	172
Thalamus	16, -18, -2	3.40	52
	-2, -4, -4	3.26	27
Middle frontal gyrus	34, 0, 62	3.38	27
Anterior insular cortex	36, 18, 0	3.40	65
Frontal pole	42, 42, 10	3.19	95
Caudate nucleus	-18, 28, 6	3.15	138
	10, 16, -2	2.80	6

^aBoth auditory stimuli and motor output were subtracted out of the brain activity, but activity in the motor cortices and cerebellum remains. MNI coordinates ($p < 0.05$ FDR-corrected, minimum spatial extent threshold of 5 voxels) are shown as well as the Z score and spatial extent in voxels (see also Fig. 4B).

*Significant after clusterwise FWE correction ($p < 0.05$).

cantly higher activations in the active-acoustic condition were found in the vocal motor centers of the primary motor cortex and in the cerebellum (Fig. 4B; Table 1). This is not surprising because the active acoustic condition includes a motor component, the clicking, whereas the passive-acoustic condition does not. However, these activation differences persist, although the active null condition was subtracted before the active-minus-passive subtraction. In particular, the precentral and postcentral gyri were active, with the peak voxel $z = \sim 27$ mm, the cerebellar vermis VI was active bilaterally, and smaller activations in the frontal regions, the anterior insula, the thalamus, caudate nucleus, and precuneus were found. The reverse comparison showed no significantly stronger activations in the passive-acoustic condition than in the active acoustic condition.

Active echolocation only

The results of the active versus passive echolocation experiment suggest that active echolocation improves performance and in-

creases brain activity in motor centers, although the output-related motor components were subtracted from the analysis. However, in that experiment, subjects were required to switch between active call production and passive listening, which may have led to activity more related to task switching than to the actual task (Dove et al., 2000). We therefore performed an additional fMRI experiment where participants only performed active echolocation. In addition to characterizing the activity during active echolocation, we examined the effect of the stimulus factors BRIR compression factor and amplitude changes on performance and on brain activity.

Behavioral performance

Subjects’ performance was similar to the previous experiment (Fig. 5A). The spectral and temporal properties of the clicks produced were consistent across conditions within subjects. Both BRIR compression factor (repeated-measures ANOVA, $F_{(3,30)} = 488.34, p = 0$) and BRIR amplitude ($F_{(3,30)} = 39.64, p = 1.47 \times 10^{-10}$) statistically affected room size rating, and the two factors showed a significant interaction ($F_{(9,90)} = 2.45, p = 0.015$). All BRIR compression factors were rated significantly different from one another. To some extent, the subjects’ ratings also reflected the small changes in BRIR amplitude. The larger the BRIR compression factor, the more different the rating was from the ratings of neighboring BRIR amplitudes. Specifically, when the BRIR was compressed by a factor of 0.2, corresponding to the smallest room, ratings ranged between 1 and 1.6 on the 1–10 scale. For a compression factor of 1, ratings ranged between 7.9 and 8.9.

Although we cannot assume a linear relationship between the stimulus parameters and the rating responses, the ratings more accurately reflect changes in BRIR compression than sound level changes induced by compression, independent of the amplitude changes that were introduced. The physical BRIR sound level increases by 5 dB when the compression factor is increased from 0.2 to 0.5, but the sound level increases by only 2 dB when the compression factor increased from 0.7 to 1. However, the subjects’ ratings changed the same amount from 0.2 to 0.5 as from 0.7 to 1, the same amount as the relative change in BRIR compression factor. This suggests that subjects relied more on stimulus factors

directly related to the BRIR compression factor, such as reverberation time, to estimate the perceived room size. Loudness and other factors not controlled for in this study may play a more important role in echolocation under different circumstances (Kolarik et al., 2014).

Motor activity patterns during active sensing

In the neuroimaging analyses, we were interested in the brain regions with a higher hemodynamic signal during all echolocation conditions (across all BRIR amplitude and compression factors) compared with the null condition. This means that the sensory information was very different between the conditions tested, but the motor components were the same. First, we compared the active versus active null conditions from the active versus passive experiment to the active versus active null conditions in this experiment using a two-sample *t* test. The differential brain activation patterns did not significantly differ between these two experiments. The activity patterns that we find for active echolocation in this experiment are likely generalizable to the passive versus active echolocation experiment.

We then examined the brain activity patterns that were higher during active echolocation than when subjects vocalized but did not receive auditory feedback (active null). The common pattern of activity across subjects included primary and higher-level auditory processing centers (Fig. 5B; for anatomical locations, see Table 2), which is to be expected as more auditory information was present during echolocation than during the null condition. Surprisingly, however, both motor and premotor centers, together with the basal ganglia and parts of the cerebellum, were significantly more active during echolocation with auditory feedback than without. These data clearly show that variation of sensory feedback can modulate vocal-motor brain activity, although vocal-motor output is unchanged.

It is reasonable to suggest that sensory differences in this echolocation paradigm involve sensory-motor coupling (Wolpert et al., 1995), thereby reflecting the active nature of echolocation. Indeed, the activity in the primary and premotor areas cannot be explained by varying motor output because the number of clicks per trial and their loudness did not differ between the active echolocation and the active null conditions.

Brain activity related to perceived room size

Another important question is whether the stimulus parameters and reported room sizes are reflected in the brain activity on a trial-by-trial basis. In fMRI, a parametric analysis identifies voxels whose BOLD response covaries with an experimental parameter. An example peak voxel in a parametric analysis from one subject is shown in Figure 6. The BOLD response of this voxel, located in the supramarginal gyrus of the parietal cortex (MNI coordinates *x, y, z* = 57, -27, 45), is plotted as a function of the three stimulus parameters. The BOLD response increases significantly with increases of either the rated room size or the BRIR compression factor, and does not change significantly with BRIR amplitude (compare Fig. 6).

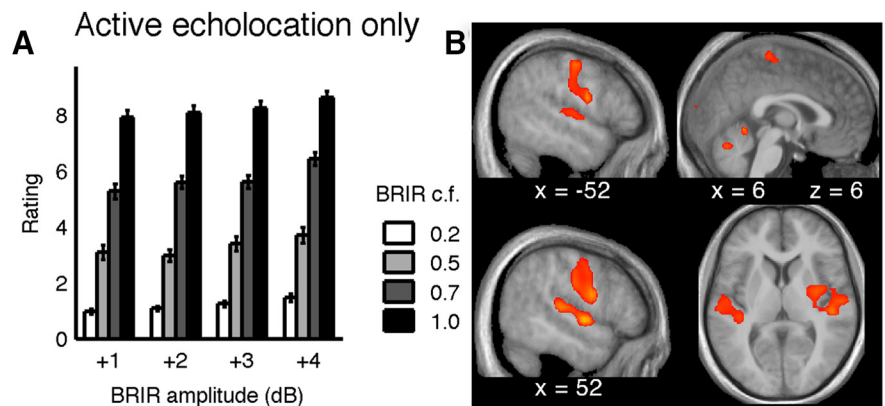


Figure 5. Active echolocation only. **A**, The room size rating is shown for the four different BRIR compression factors and as a function of BRIR amplitude. Error bars indicate SE across subjects. The data show that, whereas the BRIR compression factor is strongly reflected in the subjects' classifications, the BRIR amplitude has a much smaller effect on the perceived room size. **B**, Regions of activity during active echolocation (active sound production with auditory feedback) compared with sound production without feedback. The auditory cortex was active bilaterally as well as primary motor areas, cerebellum, and the visual pole (for details, see Table 2). Activity maps were thresholded at *p* < 0.05 (FDR corrected) and overlaid on the mean structural image of all subjects in the study. *x* and *z* values indicate MNI coordinates of the current slice.

Table 2. Spatial coordinates of the local hemodynamic activity maxima during echolocation versus null (click production without auditory feedback)^a

Region	<i>x, y, z</i> (mm)	Z score	Extent
Subcortical			
Thalamus, premotor	-15, -18, 9	3.06	5
Cortical auditory			
Temporal pole*	54, -6, -3	4.67	1126
Heschl's gyrus (H1, H2)	48, -24, 9	4.09	
Superior temporal lobe	-39, -27, 0	3.45	181
Heschl's gyrus (H1, H2)	-51, -12, 3	3.24	
Planum temporale	-60, -21, 6	3.29	
Cortical sensorimotor, frontal			
Precentral gyrus*	-48, -3, 18	4.48	970
Precentral gyrus, BA6	-63, 0, 18	4.29	
Middle cingulate cortex	-9, -3, 30	4.34	
Juxtapositional cortex, BA6	-3, -6, 51	3.08	11
Precentral gyrus ζ	54, 3, 21	4.26	
Cortical visual			
Occipital pole	3, -93, 24	3.11	8
Cerebellum			
Right I–V	3, -51, -6	3.98	18
Right VI, Crus I	24, -66, -21	3.60	18
Vermis VI	3, -66, -21	3.57	38

^aAll coordinates are from the group analysis (*p* < 0.05 FDR-corrected, spatial extent threshold of 5 voxels) given in MNI space, as well as the Z score and the cluster extent size (see also Fig. 5). Coordinates without extent values are subclusters belonging to the next closest cluster.

*Significant after clusterwise FWE correction (*p* < 0.05). ζ belongs to cluster 54, -6, -3.

Using a single-subject statistical model that included room size rating as well as BRIR amplitude variations, we identified brain regions where the BOLD response was significantly and positively correlated with the rated room size (Fig. 7; Table 3). In line with the findings from the subtractive analysis (Fig. 5B), activation in both auditory cortices and cortical motor areas were found (Fig. 7). This strengthens the conclusion that activations in sensory and motor cortices are tightly coupled during active echolocation.

In addition to cortical auditory and motor regions, activity in the medial geniculate nucleus (MGN) and the inferior colliculus (IC) was correlated with room size rating. These areas are well-described subcortical auditory-sensory nuclei. Activity in these

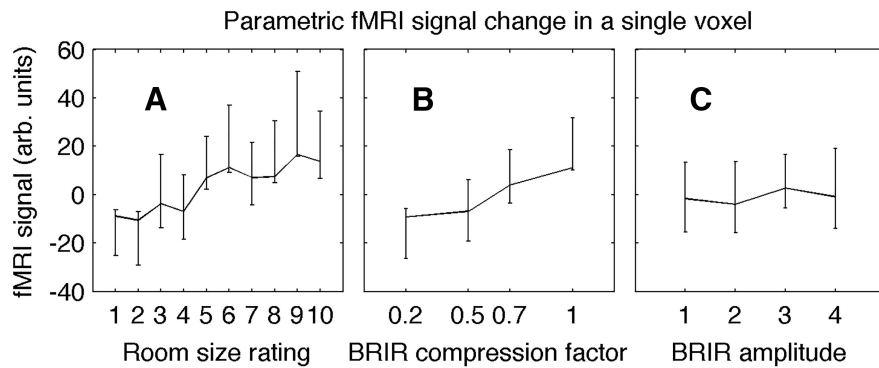


Figure 6. An example of the parametric modulations in the hemodynamic response with respect to the experimental parameters. The BOLD signal values in a single voxel (MNI coordinates $x, y, z = 57, -27, 45$) in the right supramarginal gyrus of the inferior parietal lobe were averaged over room size rating (A), reverberation scaling (B), and amplitude (C) in an example subject. It is clear here that activity in this voxel was related to both the reverberation scaling and room size rating but not the amplitude. All three experimental parameters were used to model activity across the brain (Fig. 7). Data are mean \pm SEM.

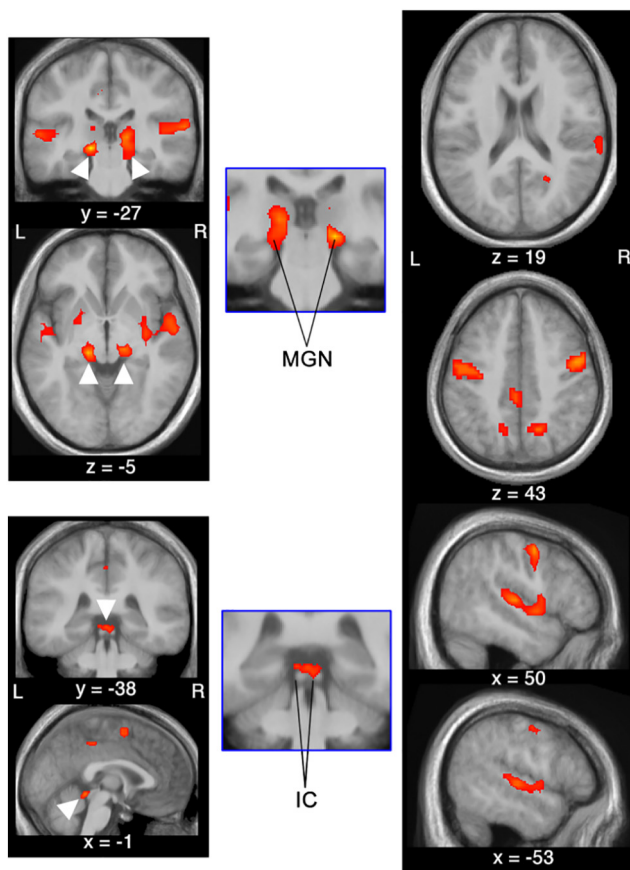


Figure 7. Areas of activity that were significantly linearly modulated by room size rating. Interestingly, both the medial geniculate nucleus (MGN) and the inferior colliculus (IC) were modulated by room size rating but not by amplitude. In addition to primary auditory centers, visual cortical areas and vocal-motor areas were also modulated by room size. The parametric vocal-motor activation is especially intriguing because the vocal-motor output does not vary with perceived room size, but still the motor-cortical activation does. Activity maps were thresholded at $p < 0.05$ (FDR corrected) and overlaid on the mean structural image of all subjects in the study. x, y, z values indicate MNI coordinates of the current slice.

areas may be driven either directly by the sensory input or by cortical feedback loops (Bajo et al., 2010). The fact that the activations significantly covaried with the rated room size but not with BRIR amplitude points toward an involvement of feedback loops. Indeed, we compared the results of the model with room

size rating and BRIR amplitude variations, with a model with BRIR compression factor and amplitude variations and found that the activity in the MGN and IC was not significantly correlated with BRIR compression factor. This supports the proposal that the subcortical activity found is related to cognition, rather than sensory input.

Finally, parametric activations were seen in the parietal and occipital cortex. These activations may be due to visual imagery (Cavanna and Trimble, 2006) and/or a modality-independent representation of space (Weeks et al., 2000; Kupers et al., 2010). Because we find parietal and occipital cortex activity in most of our analyses, it is not possible to differentiate whether the activity is more linked to the

perceived space than to the presence of auditory sensory information in general.

BOLD signal activity did not significantly covary with BRIR amplitude in any voxel in the brain, even at the less conservative threshold of $p < 0.001$ uncorrected for multiple comparisons, and 0 voxel threshold. This lenient threshold provides a better control of false negatives, but still no significant covariation of brain activity with BRIR amplitude was found. This supports the behavioral evidence that our subjects were judging room size based on BRIR compression factor more than on BRIR amplitude. However, with this design, we cannot separate out what component of the BRIR compression subjects used to solve the task.

Brain activity in a blind echolocation expert

To examine the brain regions involved in active sensing when echolocation has been performed from an early age, brain activity was measured from a single congenitally blind echolocation expert engaged in the room size estimation task with active echolocation. Since his childhood, this subject has gathered information about his surroundings by producing tongue clicks and listening to how the clicks bounce back from objects around him.

Despite lack of previous training on the psychophysical paradigm, the blind echolocation expert solved the psychophysical task in the scanner very well. His ratings of perceived room size were very similar to those of the (extensively trained) sighted subjects (Fig. 8A compared with Fig. 5A). Results from a subtractive analysis for this single blind subject are shown in Figure 8B and Table 4 in the same format as for the sighted subjects in Figure 5B. This blind subject did not show activation in primary auditory areas but strong and extended activations in primary-visual areas (right occipital cortex). This confirms earlier reports showing activity in primary visual areas during auditory and tactile tasks in the early blind (Kupers et al., 2010). In particular, we found activity in the middle occipital gyrus, which is known to be specialized for spatial processing tasks in the early blind (Renier et al., 2010). The only active auditory area was the left planum temporale, a part of auditory cortex involved in the processing of spatial auditory information (Griffiths and Warren, 2002). Strong activations are seen in (mostly right) parietal cortex. These activations partially overlap with the parietal parametric activations found in the sighted subjects (compare Fig. 7).

The activation pattern seen with the current experimental paradigm is qualitatively similar to the activity seen in an early

Table 3. Spatial coordinates of the local hemodynamic activity maxima for the linear correlation with subjective room size rating^a

Region	x, y, z (mm)	Z score	Extent
Subcortical			
Medial geniculate body*	−15, −27, −6	5.37	1634
Inferior colliculus	0, −42, −9	3.95	
Thalamus			
Premotor, prefrontal	−12, −15, −3	3.46	
Acoustic radiation	15, −24, −3	4.77	
Acoustic radiation	−12, −33, 6	3.57	
Corticospinal tract	−15, −24, 12	3.51	
Pallidum	−18, −3, 0	3.79	
Putamen	−33, −18, −6	3.22	
Cortical auditory			
Heschl's gyrus (H1, H2)			
Planum temporale	57, −15, 6	4.62	
Superior temporal lobe*	−57, −24, 3	4.53	267
Planum polare	−51, −3, 0	3.77	
Cortical sensorimotor, frontal			
Primary somatosensory cortex, BA3a	42, −6, 30	3.58	
Superior frontal gyrus, BA6	−27, −6, 60	3.82	
Superior frontal gyrus, BA6	−9, 6, 69	3.49	
Middle frontal gyrus	−51, 9, 48	3.48	
Premotor cortex, BA6*	9, 0, 54	3.97	188
Precentral gyrus, BA6*	51, −3, 48	4.97	136
Precentral gyrus BA4a*	−42, −12, 51	4.32	268
Anterior insular cortex	−36, 12, −12	3.84	
Cortical visual, parietal			
Calcarine sulcus	21, −57, 21	3.37	9
Precuneus	−15, −63, 51	3.74	48
	15, −60, 42	4.06	46
Posterior cingulate gyrus	−3, −33, 45	3.46	27
	−9, −27, 39	3.21	

^aAll coordinates are from the group analysis ($p < 0.05$ FDR-corrected, spatial extent threshold of 5 voxels) given in MNI space, as well as the Z score and the cluster extent size (see also Fig. 8). Coordinates without extent values are subclusters belonging to the next closest cluster.

*Significant after clusterwise FWE correction ($p < 0.05$).

blind subject in a passive echolocation task compared with silence (Thaler et al., 2011), in particular the lack of auditory activity when comparing the presence or absence of echoes. More detailed comparisons of the two studies are difficult, however, because the relative difference in auditory information between the task and control conditions in the two studies was very different. Interestingly, we see very little activity in the cerebellum and primary motor cortex (Table 4) in our active echolocation task. The motor activity was instead seen in the parametric modulation with room size. Although otherwise instructed, the current echolocation expert adjusted both emission loudness and repetition frequency based on the perceived room size. Although this strategy is perceptually useful, as evidenced from echolocating species of bats and toothed whales, it confounds the intended sensory-evoked parametric analysis. Any parametric modulation of brain activity with room size in the echolocation expert can be a result of both sensory and motor effects. Thus, the behavioral strategy of the echolocation expert precludes quantification of the selective modulation of brain activity by sensory input.

To quantify the differences in brain activity between the subject groups, we used a two-sample group-level GLM to test the differences between the blind subject and the sighted subjects. The pattern of brain activity seen in the single analysis for the blind subject was significantly higher than in sighted individuals. However, no regions of the brain showed significantly higher activity for the sighted subjects, suggesting that, in the blind echolocation expert, subthreshold motor activity was still present during active echolocation compared with the active null.

Discussion

Echolocation is a unique implementation of active sensing that probes the spatial layout of the environment without vision. Using a virtual echo-acoustic space technique, we were able to explore the production of vocalizations and the auditory analysis of their echoes, in a fully controlled, rigid paradigm. The current psychophysical results demonstrate that sighted humans can be effectively trained to discriminate changes in the size of an acoustically excited virtual space with an acuity comparable with visual spatial-frequency discrimination (Greenlee et al., 1990). To solve this task, subjects excited the virtual room by producing a series of short, loud, and broadband vocalizations (typically tongue clicks). As would be expected in active sensing, the psychophysical performance was related to the vocalizations produced. Subjects that produced fewer but louder clicks performed better (Fig. 3).

Echo-acoustic room size discrimination in humans has previously only been characterized qualitatively. McCarthy (1954) described a blind echolocating child who “entered a strange house, clicked once or twice and announced that it was a large room.” McGrath et al. (1999) showed that, using echoes from their own voices, humans can discriminate a small room with a size of 3 m × 3 m × 2.5 m from a concert hall with the dimensions of 60 m × 80 m × 20 m. Quantitative information does exist about the passive evaluation of the reverberation times of rooms. When presented with synthetic BRIRs consisting of temporally decaying bands of noise, subjects' JNDs are between 5% and 10% of the reference reverberation time (Seraphim, 1958). Our subjects were similarly good at estimating changes in room size, but based on the auditory analysis of active, self-generated vocalizations. Passively presenting the BRIRs themselves, instead of convolving them with a source sound, provides an auditory stimulation that approaches a Dirac Impulse (like a slash from a whip). Self-generated vocalizations are not as broadband as synthesized BRIRs, and there is increased masking of the source onto the reverberation. However, in our experiment, active vocalization led to better room size classification performance than passive listening (Fig. 4), supporting the idea that additional and perhaps redundant information, in this case from the motor system, increases performance (Schenkman and Nilsson, 2010).

The evaluation of room size based on the evaluation of reverberation from self-generated sounds may involve the estimation of egocentric distance from sound-reflecting surfaces. Perception of reverberation and its application for echo-acoustic orientation are comprehensive (for review, see Kaplanis et al., 2014; and Kolarik et al., 2016, respectively). For instance, the direct-to-reverberant ratio of an external sound reliably encodes the distance of its source, and changes thereof encode changes in source distance (Bronkhorst and Houtgast, 1999). Zahorik (2002) found, however, that psychophysical sensitivity to changes of the direct-to-reverberant ratio is on the order of 5–6 dB, corresponding to an approximately twofold change in the egocentric distance toward a sound source. This ratio is too large to explain the high sensitivity to changes in room size that we have shown here. Instead, current psychophysical performance is more likely to be governed by evaluation of changes in reverberation time (Seraphim, 1958), supported also by the relatively low degree of sensitivity to BRIR amplitude changes in the room size estimation experiment (Fig. 5A). Reverberation time, together with interaural coherence, is the main perceptual cue used to assess room acoustics (Hameed et al., 2004; Zahorik, 2009). Only Cabrera et al. (2006) have indicated that perceptual clarity of reproduced

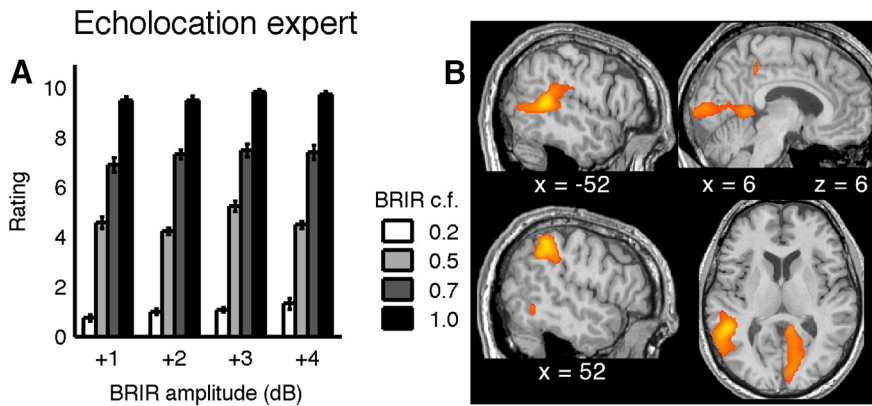


Figure 8. Blind echolocation expert. *A*, Psychophysical performance as in Figure 5*A*. Bars represent the mean room size classification as a function of BRIR compression factor (grayscale of the bars) and BRIR amplitude (bar groups). Error bars indicate SEs across trial repetitions. Without any prior training, the classification is very stable and similar to that of the extensively trained, sighted subjects. *B*, Regions of activity in an echolocation expert during active echolocation compared with sound production without auditory feedback. The strongest regions of activations in the fMRI data were found in visual and parietal areas (compare Table 4). Activity maps were thresholded at $p < 0.05$ (FDR corrected) and overlaid on the subject’s normalized structural image.

Table 4. Spatial coordinates of the local hemodynamic activity maxima during echolocation versus null (click production without auditory feedback) in an echolocation expert^a

Region	<i>x, y, z</i> (mm)	Zscore	Extent
Inferior parietal cortex, supramarginal gyrus*	48, -44, 42	5.07	1027
Temporal-parietal-occipital junction*	-52, -46, 8	4.69	2162
Fusiform gyrus	32, -60, -12	4.30	424
Calcarine cortex, cuneus, posterior cingulum, V1*	16, -54, 18	4.21	1722
Cerebellum Crus 1	-38, -58, -34	4.06	244
Precuneus and posterior cingulum	8, -40, 42	3.49	128
Inferior temporal gyrus, bordering occipital cortex	50, -54, -8	3.27	49
Paracentral lobule	-14, -26, 72	3.42	53
Occipital pole	22, -90, 32	3.22	23
Cerebellum, Crus II	-16, -72, -36	3.03	14
Precentral gyrus	-56, -10, 40	3.03	18
Cuneus, V2	8, -88, 24	2.99	7

^aMNI coordinates ($p < 0.05$ FDR-corrected, spatial extent threshold of 5 voxels), together with the Zscore, and the cluster extent are shown (see also Fig. 8*B*).

*Significant after clusterwise FWE correction ($p < 0.05$).

speech sounds may carry even greater information about room size than reverberation time.

Although in our paradigm active echolocation improves performance over passive echolocation, assisted or passive echolocation may be more useful in other circumstances. The sensory-motor coupling in active echolocation requires extensive training; and even then, performance differs greatly across participants, similar to the ability to pronounce non-native speech sounds (Kartushina et al., 2015). Participants that are naive to active echolocation detect ensonified objects better when passive echolocation is used (Thaler and Castillo-Serrano, 2016). After training with multisensory sensory substitution devices using passive auditory information, navigation performance can improve to a level similar to sighted navigation, although many limitations still exist (Chebat et al., 2015).

Using the virtual echo-acoustic space, we were able to investigate brain activity while subjects are engaged in echolocation, and thereby separate out the individual sensory and motor components of human echolocation. Primary and secondary motor cortices have previously been found in both blind and sighted subjects during passive echolocation (Thaler et al., 2011), although there the activity may be a result of motor imagery, or

motor activity during action observation (Cattaneo and Rizzolatti, 2009; Massen and Prinz, 2009). In the current study, both auditory and motor cortices were more active when auditory feedback was present (Fig. 5*B*) than when subjects vocalized without auditory feedback. Primary somatosensory and motor cortex activity together with the cerebellum were significantly more active during echolocation than when one of the modalities, audition or motor control, was present without the other. These motor areas also showed activity that was correlated with the auditory percept. Together, these results provide strong evidence that motor feedback is a crucial component of echolocation.

The vast majority of animal sensory systems (also in humans) passively sample the environment (i.e., extrinsic energy sources, such as light or sound stimulate sensory receptors). Still, animals generally use the motor system to sample the environment (e.g., to focus the eyes or turn the ears), but truly active senses, where the animal itself produces the energy used to probe the surroundings, are rare in the animal kingdom (Nelson and Maciver, 2006). Examples comprise the active electric sense by weakly electric fishes (Lissmann and Machin, 1958) and echolocation, where sensing of the environment occurs through auditory analysis of self-generated sounds (Griffin, 1974). The advanced echolocation systems of bats and toothed whales involve dynamic adaptation of the outgoing sound and behavior (e.g., head aim and flight path) based on perception of the surroundings through auditory processing of the information carried by returning echoes.

The motor system can modulate sensory information processing, independent of whether the energy sensed is also produced. Temporal motor sequences, or rhythmic movements, sharpen the temporal auditory stimulus selection through top-down attentional control (Morillon et al., 2014). Motor output is regulated in part by slow motor cortical oscillatory rhythms that have also been shown to affect the excitability of task-relevant sensory neurons (Schroeder et al., 2010). Our results support this idea in a classical active sensing task. If the temporal comparison between call and reverberation is used in evaluating room size, as it appears to be, then this may be a possible neural mechanism that would explain both our behavioral and neuroimaging results.

In addition to the motor system, active echolocation recruited cortical and subcortical auditory processing regions, as well as visual and parietal areas not typically known for auditory processing. As in the visual cortex, the auditory cortex is thought to comprise two processing streams: the dorsal or “where” stream, and the ventral or “what” stream (Rauschecker and Tian, 2000). Sound localization and spatial hearing recruit early auditory areas posterior and lateral to the primary auditory cortex, extending into the parietal cortex both in humans and nonhuman primates (Rauschecker and Tian, 2000; Alain et al., 2001; van der Zwaag et al., 2011). Recently, the function of the dorsal auditory stream was reconceptualized to involve sensory-motor control and integration in speech (Rauschecker, 2011; Chevillet et al., 2013). Although our experimental paradigm involved spatial auditory processing (classically the “where” stream), the vocal-motor requirements of human echolocation also challenge

sensory-motor integration, making it conceivable with a non-spatial task to further delineate auditory processing streams.

Auditory midbrain (IC) and thalamus (MGN) activity was modulated by the behavioral output variable on a trial-by-trial basis. Both the IC and MGN are part of the ascending auditory system, but corticocollicular feedback was shown to play a crucial role in auditory spatial learning and plasticity (Bajo et al., 2010). Based on our results, corticocollicular feedback may also contribute to sonar processing.

Both the sighted subjects and the blind echolocation expert had visual and parietal activity during echolocation. For the sighted subjects, activity in the precuneus, in the medial parietal cortex, may be a result of visual imagery (Cavanna and Trimble, 2006). Sighted persons typically visualize nonvisual tasks, and visual imagery is positively correlated with echolocation performance (Thaler et al., 2014a). Alternatively, the parietal activity may reflect a modality-independent representation of space. Auditory localization activated the medial parietal areas, including the precuneus in both sighted and blind subjects (Weeks et al., 2000), and is active during imagined passive locomotion without visual memory (Wutte et al., 2012). Parietal areas were active in when both blind and sighted subjects used passive echolocation for path finding (Fiehler et al., 2015). Route navigation using a tactile sensory substitution device activates the precuneus in congenitally blind subjects and in visual route navigation in sighted subjects (Kupers et al., 2010). This evidence speaks for multimodal spatial processing for action in the parietal cortex in humans.

References

- Alain C, Arnott SR, Hevenor S, Graham S, Grady CL (2001) “What” and “where” in the human auditory system. *Proc Natl Acad Sci U S A* 98:12301–12306. [CrossRef Medline](#)
- Amaro E Jr, Williams SC, Shergill SS, Fu CH, MacSweeney M, Picchioni MM, Brammer MJ, McGuire PK (2002) Acoustic noise and functional magnetic resonance imaging: current strategies and future prospects. *J Magn Reson Imaging* 16:497–510. [CrossRef Medline](#)
- Ashburner J, Friston KJ (2005) Unified segmentation. *Neuroimage* 26:839–851. [CrossRef Medline](#)
- Bajo VM, Nodal FR, Moore DR, King AJ (2010) The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat Neurosci* 13:253–260. [CrossRef Medline](#)
- Blauert J (1997) Spatial hearing: the psychophysics of human sound localization: Cambridge, MA: Massachusetts Institute of Technology.
- Blauert J, Lindemann W (1986) Auditory spaciousness: some further psychoacoustic analyses. *J Acoust Soc Am* 80:533–542. [CrossRef Medline](#)
- Blauert J, Xiang N (1993) Binaural scale modelling for auralization and prediction of acoustics in auditoria. *J Appl Acoust* 38:267–290. [CrossRef](#)
- Bronkhorst AW, Houtgast T (1999) Auditory distance perception in rooms. *Nature* 397:517–520. [CrossRef Medline](#)
- Burton G (2000) The role of the sound of tapping for nonvisual judgment of gap crossability. *J Exp Psychol Hum Percept Perform* 26:900–916. [CrossRef Medline](#)
- Cabrera DP, Jeong D (2006) Auditory room size perception: a comparison of real versus binaural sound-fields. *Proceedings of Acoustics*. Christchurch, New Zealand.
- Cattaneo L, Rizzolatti G (2009) The mirror neuron system. *Arch Neurol* 66:557–560. [CrossRef Medline](#)
- Cavanna AE, Trimble MR (2006) The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129:564–583. [CrossRef Medline](#)
- Chebat DR, Maidenbaum S, Amedi A (2015) Navigation using sensory substitution in real and virtual mazes. *PLoS One* 10:e0126307. [CrossRef Medline](#)
- Chevillet MA, Jiang X, Rauschecker JP, Riesenhuber M (2013) Automatic phoneme category selectivity in the dorsal auditory stream. *J Neurosci* 33:5208–5215. [CrossRef Medline](#)
- Dove A, Pollmann S, Schubert T, Wiggins CJ, von Cramon DY (2000) Prefrontal cortex activation in task switching: an event-related fMRI study. *Brain Res Cogn Brain Res* 9:103–109. [CrossRef Medline](#)
- Fiehler K, Schütz I, Meller T, Thaler L (2015) Neural correlates of human echolocation of path direction during walking. *Multisens Res* 28:195–226. [CrossRef Medline](#)
- Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15:870–878. [CrossRef Medline](#)
- Greenlee MW, Gerling J, Waltenspiel S (1990) Spatial-frequency discrimination of drifting gratings. *Vision Res* 30:1331–1339. [CrossRef Medline](#)
- Griffin DR (1974) Listening in the dark: acoustic orientation of bats and men. Mineola, NY: Dover.
- Griffiths TD, Warren JD (2002) The planum temporale as a computational hub. *Trends Neurosci* 25:348–353. [CrossRef Medline](#)
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) “Sparse” temporal sampling in auditory fMRI. *Hum Brain Mapp* 7:213–223. [CrossRef Medline](#)
- Hameed SP, Valde K, Pulkki V (2004) Psychoacoustic cues in room size perception. In: *Audio Engineering Society Convention*, 116. Berlin, Germany.
- Hidaka T, Beranek LL (2000) Objective and subjective evaluations of twenty-three opera houses in Europe, Japan, and the Americas. *J Acoust Soc Am* 107:368–383. [CrossRef Medline](#)
- Kaplanis NB, Jensen SH, van Waaterschoot T (2014) Perception of reverberation in small rooms: a literature study. *Proceedings of the AES 55th Conference on Spatial Audio*. Helsinki, Finland.
- Kartushina N, Hervais-Adelman A, Frauenfelder UH, Golestani N (2015) The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *J Acoust Soc Am* 138:817–832. [CrossRef Medline](#)
- Kolarik AJ, Cirstea S, Pardhan S, Moore BC (2014) A summary of research investigating echolocation abilities of blind and sighted humans. *Hear Res* 310:60–68. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Kupers R, Chebat DR, Madsen KH, Paulson OB, Ptito M (2010) Neural correlates of virtual route recognition in congenital blindness. *Proc Natl Acad Sci U S A* 107:12716–12721. [CrossRef Medline](#)
- Lissmann HW, Machin KE (1958) The mechanism of object location in *Gymnarchus niloticus* and similar fish. *J Exp Biol* 35:451–486. [CrossRef](#)
- Litovsky RY, Colburn HS, Yost WA, Guzman SJ (1999) The precedence effect. *J Acoust Soc Am* 106:1633–1654. [CrossRef Medline](#)
- Massen C, Prinz W (2009) Movements, actions and tool-use actions: an ideomotor approach to imitation. *Philos Trans R Soc Lond B Biol Sci* 364:2349–2358. [CrossRef Medline](#)
- McCarthy BW (1954) Rate of motion and object perception in the blind. *New Outlook for the Blind* 48:316–322.
- McGrath R, Waldmann T, Fernström M (1999) Listening to rooms and objects. In: *16th International Conference: Spatial Sound Reproduction*. Rovaniemi, Finland.
- Morillon B, Schroeder CE, Wyart V (2014) Motor contributions to the temporal precision of auditory attention. *Nat Commun* 5:2555. [CrossRef Medline](#)
- Nelson ME, MacIver MA (2006) Sensory acquisition in active sensing systems. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 192:573–586. [CrossRef Medline](#)
- Nielsen JB, Dau T (2010) Revisiting perceptual compensation for effects of reverberation in speech identification. *J Acoust Soc Am* 128:3088–3094. [CrossRef Medline](#)
- Rauschecker JP (2011) An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear Res* 271:16–25. [CrossRef Medline](#)
- Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci U S A* 97:11800–11806. [CrossRef Medline](#)
- Renier LA, Anurova I, De Volder AG, Carlson S, VanMeter J, Rauschecker JP (2010) Preserved functional specialization for spatial processing in the middle occipital gyrus of the early blind. *Neuron* 68:138–148. [CrossRef Medline](#)

- Rice CE (1967) Human echo perception. *Science* 155:656–664. [CrossRef Medline](#)
- Rojas JA, Hermosilla JA, Montero RS, Espí PL (2009) Physical analysis of several organic signals for human echolocation: oral vacuum pulses. *Acta Acustica Acustica* 95:325–330. [CrossRef](#)
- Sabine WC (1923) *Collected Papers on Acoustics*. Cambridge: Harvard University Press.
- Schenkman BN, Nilsson ME (2010) Human echolocation: blind and sighted persons' ability to detect sounds recorded in the presence of a reflecting object. *Perception* 39:483–501. [CrossRef Medline](#)
- Schörnich S, Wallmeier L, Gesele N, Nagy A, Schraner M, Kish D, Wiegrebe L (2013) Psychophysics of human echolocation. *Adv Exp Med Biol* 787:311–319. [CrossRef Medline](#)
- Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P (2010) Dynamics of active sensing and perceptual selection. *Curr Opin Neurobiol* 20:172–176. [CrossRef Medline](#)
- Schuller G, Fischer S, Schweizer H (1997) Significance of the paralemniscal tegmental area for audio-motor control in the moustached bat, *Pteronotus p. parnellii*: the afferent and efferent connections of the paralemniscal area. *Eur J Neurosci* 9:342–355. [CrossRef Medline](#)
- Seraphim HP (1958) Untersuchungen über die Unterschiedsschwelle Exponentiellen Abklingens von Rauschbandimpulsen. *Acustica* 8:280–284.
- Seraphim HP (1961) über die Wahrnehmbarkeit mehrerer Rückwürfe von Sprachschall. *Acustica* 11:80–91.
- Smotherman MS (2007) Sensory feedback control of mammalian vocalizations. *Behav Brain Res* 182:315–326. [CrossRef Medline](#)
- Stoffregen TA, Pittenger JB (1995) Human echolocation as a basic form of perception and action. *Ecol Psychol* 7:181–216. [CrossRef](#)
- Teng S, Whitney D (2011) The acuity of echolocation: spatial resolution in the sighted compared to expert performance. *J Vis Impair Blind* 105:20–32. [Medline](#)
- Thaler L, Castillo-Serrano J (2016) People's ability to detect objects using click-based echolocation: a direct comparison between mouth-clicks and clicks made by a loudspeaker. *PLoS One* 11:e0154868. [CrossRef Medline](#)
- Thaler L, Arnott SR, Goodale MA (2011) Neural correlates of natural human echolocation in early and late blind echolocation experts. *PLoS One* 6:e20162. [CrossRef Medline](#)
- Thaler L, Wilson RC, Gee BK (2014a) Correlation between vividness of visual imagery and echolocation ability in sighted, echo-naïve people. *Exp Brain Res* 232:1915–1925. [CrossRef Medline](#)
- Thaler L, Milne JL, Arnott SR, Kish D, Goodale MA (2014b) Neural correlates of motion processing through echolocation, source hearing, and vision in blind echolocation experts and sighted echolocation novices. *J Neurophysiol* 111:112–127. [CrossRef Medline](#)
- Thomas JA, Moss CF, Vater M (2004) *Echolocation in bats and dolphins*. Chicago: University of Chicago.
- van der Zwaag W, Gentile G, Gruetter R, Spierer L, Clarke S (2011) Where sound position influences sound object representations: a 7-T fMRI study. *Neuroimage* 54:1803–1811. [CrossRef Medline](#)
- Wallmeier L, Kish D, Wiegrebe L, Flanagin VL (2015) Aural localization of silent objects by active human biosonar: neural representations of virtual echo-acoustic space. *Eur J Neurosci* 41:533–545. [CrossRef Medline](#)
- Watkins AJ (2005) Perceptual compensation for effects of reverberation in speech identification. *J Acoust Soc Am* 118:249–262. [CrossRef Medline](#)
- Watkins AJ, Makin SJ (2007) Steady-spectrum contexts and perceptual compensation for reverberation in speech identification. *J Acoust Soc Am* 121:257–266. [CrossRef Medline](#)
- Weeks R, Horwitz B, Aziz-Sultan A, Tian B, Wessinger CM, Cohen L, Hallett M, Rauschecker JP (2000) A positron emission tomographic study of auditory localization in the congenitally blind. *J Neurosci* 20:2664–2672. [Medline](#)
- Wolpert DM, Ghahramani Z, Jordan MI (1995) An internal model for sensorimotor integration. *Science* 269:1880–1882. [CrossRef Medline](#)
- Wutte MG, Glasauer S, Jahn K, Flanagin VL (2012) Moving and being moved: differences in cerebral activation during recollection of whole-body motion. *Behav Brain Res* 227:21–29. [CrossRef Medline](#)
- Zahorik P (2002) Direct-to-reverberant energy ratio sensitivity. *J Acoust Soc Am* 112:2110–2117. [CrossRef Medline](#)
- Zahorik P (2009) Perceptually relevant parameters for virtual listening simulation of small room acoustics. *J Acoust Soc Am* 126:776–791. [CrossRef Medline](#)