Human exploration of enclosed spaces through echolocation

Virginia L. Flanagin1,3, Sven Schörnich2,3, Michael Schranner2, Nadine Hummel1, Ludwig Wallmeier3, Magnus Wahlberg4, Thomas Stephan1,5 and Lutz Wiegrebe2

1 German Center for Vertigo and Balance Disorders
2 Division of Neurobiology, Department of Biology II
3 These authors contributed equally to the presented work
4 Fjord&Bælt and University of Southern Denmark, DK-5300 Kerteminde, Denmark
5 Department of Neurology, Ludwig Maximilians University, Munich, Germany

DOI: 10.1523/JNEUROSCI.1566-12.2016
Received: 30 March 2012
Revised: 23 November 2016
Accepted: 1 December 2016
Published: 10 January 2017


Conflict of Interest: The authors declare no conflicts of interest for this study.

This study was funded by the Deutsche Forschungsgemeinschaft Wi 1518/9 to Lutz Wiegrebe and the German Center for Vertigo and Balance Disorders (BMBF IFB 01EO0901 to Virginia Flanagin). We would like to thank Daniel Kish for insightful discussions together with the organization ‘World Access for the Blind’ and Benedikt Grothe for providing exceptionally good research infrastructure and thoughtful discussions on the topic.

Corresponding author: Virginia L. Flanagin, Neurological Research Pavilion, University Hospital Munich -- Großhadern, Feodor-Lynen-Str. 19, 81377 Munich, Germany, +49 89 7095 4819, vflanagin@lrz.uni-muenchen.de

Cite as: J. Neurosci 2017; 10.1523/JNEUROSCI.1566-12.2016
Alerts: Sign up at www.jneurosci.org/cgi/alerts to receive customized email alerts when the fully formatted version of this article is published.
Human exploration of enclosed spaces through echolocation

Virginia L. Flanagin1,3, Sven Schörnich2,3, Michael Schranner2, Nadine Hummel1, Ludwig Wallmeier2, Magnus Wahlberg4, Thomas Stephan1,5, and Lutz Wiegrebe2

1German Center for Vertigo and Balance Disorders
2Division of Neurobiology, Department of Biology II
3These authors contributed equally to the presented work
4Fjord&Bælt and University of Southern Denmark, DK-5300 Kerteminde, Denmark
5Department of Neurology, Ludwig Maximilians University, Munich, Germany

Corresponding author:
Virginia L. Flanagin, Neurological Research Pavilion, University Hospital Munich - Großhadern
Feodor-Lynen-Str. 19, 81377 Munich, Germany
+49 89 7095 4819, vflanagin@lrz.uni-muenchen.de

Manuscript properties:
Number of pages: 39
Number of figures: 8
Number of tables: 4
Number of words in Abstract: 226
Number of words in Introduction: 617
Number of words in Discussion: 1490
Number of words in Significance statement: 73

The authors declare no conflicts of interest for this study.

Acknowledgements:
This study was funded by the Deutsche Forschungsgemeinschaft Wi 1518/9 to Lutz Wiegrebe and the German Center for Vertigo and Balance Disorders (BMBF IFB 01EO0901 to Virginia Flanagin). We would like to thank Daniel Kish for insightful discussions together with the organization ‘World Access for the Blind’ and Benedikt Grothe for providing exceptionally good research infrastructure and thoughtful discussions on the topic.
Abstract

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 to 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.
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. Complementary 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, or echolocation, departs from classical spatial hearing in that the listener is also the sound source, i.e. he or she must use 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.

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; Thaler et al., 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 vocalisation 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. Then 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.

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 (Project Nrs. 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, Naerum, Denmark) 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, Cambridge, USA). The chapel was acoustically excited with a 20 s sine sweep from 200 to 20000 Hz. The sweep was created with Matlab (The MathWorks, Inc., Natick, USA); playback and recording were implemented with SoundMexPro (HörTech GmbH., Oldenburg, Germany). The frequency response of the mouth simulator was digitally equalized. The sweep was amplified (Stereo Amplifier A-109,
Pioneer Electronics, Willich, Germany) 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 approximately 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) (cf, Fig 1C). Note that the spectral center of gravity of the BRIR increases with decreasing compression factor (Fig. 1C). The co-variation 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, Oldenburg, Germany) 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 less than 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 (age 23.4 ± 2.2 yrs (mean ± SD), 4 female).

**Procedure**

The psychophysical experiments were conducted in a 1.2 m x 1.2 m x 2.2 m sound-attenuated
anechoic chamber (G+H Schallschutz GmbH, Ludwigshafen, Germany). Just-noticeable
differences (JNDs) 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 and 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 towards 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, Wedemark, Germany)* and in-ear headphones *(Etymotic Research ER-4S, Grove Village, IL, USA)*. The headset microphone was positioned at a distance of about 3 cm to the left of the subjects’ mouth. Headphones and microphone were connected to an external soundcard *(RME Fireface 400, Haimhausen, Germany)*, which was connected to the PC. A gamepad *(BigBen interactive, Bergheim, Germany)* 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 to 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 standard deviation of the thresholds across the last three runs was less than ¼ of the mean threshold. JNDs are specified by the percentage of each side of 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 (Schornich 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 standard deviation 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 Rho.

Active vs. 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 yrs (mean ± SD), 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, Wedemark, Germany), amplified (Sennheiser MO 2000 CU), converted (Motu Traveler, Cambridge, USA), 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, Bergen, Norway). 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 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 delineate 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-second echolocation interval to the start of fMRI acquisition was therefore between 10.1 and 5.7 seconds. 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 seconds.

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-minute 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, Milwaukee, USA*) using a standard 8-channel head coil. 38 contiguous transverse slices (slice thickness 3.5 mm, no gap) were acquired using a gradient echo echo-planar-imaging (EPI) sequence (TR 16.5 s,
TE 40 ms, flip angle 90 deg. Matrix 64 x 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 x 0.8 x 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 x 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, UK) for Matlab. Volumes were corrected for head motion using realignment, and spatially normalized to Montreal Neurological Institute (MNI) space through segmentation of the high-resolution MR-image (Ashburner and Friston, 2005). Images were smoothed with an 8 mm full-width at half maximum isotropic Gaussian kernel to reduce spatial noise.

Single-subject effects were tested with the general linear model (GLM). High-pass filtering (cut-off 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 (HRF). 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 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 to 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 vs. passive experiment randomly switched between active production of echolocation calls and passive listening to these calls. This task switching could have lead 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 vs. 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 vs. 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 approximately 45 min).

**Analysis**

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

FMRI analysis, including preprocessing and significance levels, was the same as in the active vs. 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 to
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 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, aged 44, 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 percent 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 grey-filled room could be discriminated from the transparent room surrounding it (cf. Fig. 2 top). These discrimination 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 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-call 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 vs. 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 one to ten, 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 (rm-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) ($F_{(1,9)}=0.015$, $p=0.91$) 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 vs. passive sensing

Because the number and loudness of clicks did not differ between 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. Significantly 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, which the passive-acoustic condition does not. Note, however, that these activation differences persist although the active null condition was subtracted before the active-minus-passive subtraction. In particular the pre and postcentral gyri were active, with the peak voxel around z = 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 vs. passive echolocation experiment suggest that active echolocation improves performance and increases 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.,
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 (rm-ANOVA, $F_{(3,30)}=488.34$, $p=0$) and BRIR amplitude ($F_{(3,30)}=39.64$, $p=1.47\times10^{-10}$) 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 to 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 vs. active null conditions from the active vs. passive experiment to the active vs. 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 vs. 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, see Table 2 for anatomical locations), 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 Fig. 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 (cf. Fig. 6).

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 and 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 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 towards an involvement of feedback loops. Indeed, we compared the results of the model with room size rating and BRIR amplitude variations, to 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 to Fig. 5A). Results from a subtractive analysis for this single blind subject are shown in Fig. 8B and Table 4 in the same format as for the sighted subjects in Fig. 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 (cf. Fig. 7).

The activation pattern seen with the current experimental paradigm is qualitatively similar to the activity seen in an early blind subject in a passive echolocation task compared to 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 were 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. While 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 general linear model 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 to 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 (VEAS) 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 to 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 and Worchel (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 x 3 m x 2.5 m from a concert hall with the dimensions of 60 m x 80 m x 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 comprehensively reviewed in 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 Houtsma, 1999)). Zahorik (2002) found, however, that psychophysical sensitivity to changes of the direct-to-reverberant ratio is on the order of five to six dB, corresponding to an about 2-fold change in the egocentric distance towards 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 inter-aural coherence, is the main perceptual cue used to assess room acoustics (Hameed, 2004; Zahorik, 2009). Only Cabrera et al. (2006) has indicated that perceptual clarity of reproduced 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 naïve 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 VEAS 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. 5B) 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 like 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, 1944). 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 non-human 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).

While 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 cortico-collicular feedback was shown to play a crucial role in auditory spatial learning and plasticity (Bajo et al., 2010). Based on our results, cortico-collicular 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 non-visual 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.
Figure captions

Figure 1. Binaural room impulse response (BRIR) of a real enclosed space. (Top) A photograph of the acoustically excited room (old St. Stephanus, Gräfelfing, Germany) with the head-and-torso simulator. (Row 2) Spectrograms of the left and right BRIRs are shown. Sound pressure level is color coded between -60 and 0 dB. (Row 3) 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 are shown. The color scale is identical to the second row.

Figure 2. Just noticeable differences (JND) in room size. (Top) The 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%. The across-subject mean is shown as the right.

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 is shown. The correlation coefficients (Spearman’s Rho) are given in the top right of each panel. The analysis shows that overall, JNDs improve with increasing call level and decrease with increasing number of calls per trial.
**Figure 4.** Active vs. 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 represent standard error 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 minus active null condition) minus (passive echolocation minus 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 (see Table 1 for details).

**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 represent standard error across subjects. The data show that while 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 to sound production without feedback. The auditory cortex was active bilaterally as well as primary motor areas, cerebellum and the visual pole (see Table 2 for details). 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 refer to MNI coordinates of the current slice.

**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 (see Fig. 7). Means and standard error of the mean are
shown here.

Figure 7. Areas of activity that were significantly linearly modulated by room size rating.
Interestingly, both the MGN and the inferior colliculus 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
represent MNI coordinates of the current slice.

Figure 8. Blind echolocation expert. (A) Psychophysical performance as in Figure 5A. Bars
show the mean room-size classification as a function of BRIR compression factor (grey scale
of the bars) and BRIR amplitude (bar groups). Error bars represent standard errors 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 to sound production without auditory feedback. The
strongest regions of activations in the fMRI data were found in visual and parietal areas (cf.
Table 4). Activity maps were thresholded at \( p<0.05 \) (FDR corrected) and overlaid on the
subject’s normalized structural image.
Table 1. Spatial coordinates of the local hemodynamic activity maxima for active echolocation – the active null, without auditory feedback, vs passive echolocation compared to the baseline null condition. In other words, both 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)

<table>
<thead>
<tr>
<th>Region</th>
<th>[x,y,z] in mm</th>
<th>Z-score</th>
<th>Extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerebellum vermis V*</td>
<td>[16, -64, -20]</td>
<td>7.14</td>
<td>4634</td>
</tr>
<tr>
<td></td>
<td>[-20, -64, -20]</td>
<td>6.69</td>
<td></td>
</tr>
<tr>
<td>Postcentral gyrus, somatosensory cortex*</td>
<td>[-56, -12, 26]</td>
<td>6.83</td>
<td>4769</td>
</tr>
<tr>
<td>Precentral gyrus*</td>
<td>[60, 2, 28]</td>
<td>6.44</td>
<td>4159</td>
</tr>
<tr>
<td>Precuneus</td>
<td>[-4, -40, -50]</td>
<td>4.74</td>
<td>172</td>
</tr>
<tr>
<td>Thalamus</td>
<td>[16, -18, -2]</td>
<td>3.40</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>[-2, -4, -4]</td>
<td>3.26</td>
<td>27</td>
</tr>
<tr>
<td>Middle frontal gyrus</td>
<td>[34, 0, 62]</td>
<td>3.38</td>
<td>27</td>
</tr>
<tr>
<td>Anterior insular cortex</td>
<td>[36, 18, 0]</td>
<td>3.40</td>
<td>65</td>
</tr>
<tr>
<td>Frontal pole</td>
<td>[42, 42, 10]</td>
<td>3.19</td>
<td>95</td>
</tr>
<tr>
<td>Caudate nucleus</td>
<td>[-18, 28, 6]</td>
<td>3.15</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>[10, 16, -2]</td>
<td>2.80</td>
<td>6</td>
</tr>
</tbody>
</table>
Table 2. Spatial coordinates of the local hemodynamic activity maxima during echolocation versus null (click production without auditory feedback). All 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]

<table>
<thead>
<tr>
<th>Region</th>
<th>[x,y,z] in mm</th>
<th>Z-score</th>
<th>Extent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subcortical</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thalamus, premotor</td>
<td>[-15, -18, 9]</td>
<td>3.06</td>
<td>5</td>
</tr>
<tr>
<td><strong>Cortical auditory</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal pole*</td>
<td>[54, -6, -3]</td>
<td>4.67</td>
<td>1126</td>
</tr>
<tr>
<td>Heschl’s gyrus (H1, H2)</td>
<td>[48, -24, 9]</td>
<td>4.09</td>
<td></td>
</tr>
<tr>
<td>Superior temporal lobe</td>
<td>[-39, -27,0]</td>
<td>3.45</td>
<td>181</td>
</tr>
<tr>
<td>Heschl’s gyrus (H1,H2)</td>
<td>[-51, -12, 3]</td>
<td>3.24</td>
<td></td>
</tr>
<tr>
<td>Planum temporale</td>
<td>[-60, -21, 6]</td>
<td>3.29</td>
<td></td>
</tr>
<tr>
<td><strong>Cortical sensorimotor, frontal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precentral gyrus*</td>
<td>[-48, -3, 18]</td>
<td>4.48</td>
<td>970</td>
</tr>
<tr>
<td>Precentral gyrus, BA6</td>
<td>[-63, 0, 18]</td>
<td>4.29</td>
<td></td>
</tr>
<tr>
<td>Middle cingulate cortex</td>
<td>[-9, -3, 30]</td>
<td>4.34</td>
<td></td>
</tr>
<tr>
<td>Juxtapositional cortex, BA6</td>
<td>[-3, -6, 51]</td>
<td>3.08</td>
<td>11</td>
</tr>
<tr>
<td>Precentral gyrus ζ</td>
<td>[54, 3, 21]</td>
<td>4.26</td>
<td></td>
</tr>
<tr>
<td><strong>Cortical visual</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occipital Pole</td>
<td>[ 3, -93, 24]</td>
<td>3.11</td>
<td>8</td>
</tr>
<tr>
<td><strong>Cerebellum</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right I-V</td>
<td>[ 3, -51, -6]</td>
<td>3.98</td>
<td>18</td>
</tr>
<tr>
<td>Right VI, Crus I</td>
<td>[ 24, -66, -21]</td>
<td>3.60</td>
<td>18</td>
</tr>
<tr>
<td>Vermis VI</td>
<td>[ 3, -66, -21]</td>
<td>3.57</td>
<td>38</td>
</tr>
</tbody>
</table>
Table 3. Spatial coordinates of the local hemodynamic activity maxima for the linear correlation with subjective room size rating. All 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)

<table>
<thead>
<tr>
<th>Region</th>
<th>[x,y,z] in mm</th>
<th>Z-score</th>
<th>Extent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subcortical</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medial geniculate body*</td>
<td>[-15, -27, -6]</td>
<td>5.37</td>
<td>1634</td>
</tr>
<tr>
<td>Inferior colliculus</td>
<td>[0, -42, -9]</td>
<td>3.95</td>
<td></td>
</tr>
<tr>
<td>Thalamus:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- premotor, prefrontal</td>
<td>[-12, -15, -3]</td>
<td>3.46</td>
<td></td>
</tr>
<tr>
<td>- acoustic radiation</td>
<td>[15, -24, -3]</td>
<td>4.77</td>
<td></td>
</tr>
<tr>
<td>- acoustic radiation</td>
<td>[-12, -33, 6]</td>
<td>3.57</td>
<td></td>
</tr>
<tr>
<td>- corticospinal tract</td>
<td>[-15, -24, 12]</td>
<td>3.51</td>
<td></td>
</tr>
<tr>
<td>Pallidum</td>
<td>[-18, -3, 0]</td>
<td>3.79</td>
<td></td>
</tr>
<tr>
<td>Putamen</td>
<td>[-33, -18, -6]</td>
<td>3.22</td>
<td></td>
</tr>
<tr>
<td><strong>Cortical auditory</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heschl’s gyrus (H1,H2)</td>
<td>[48, -21, 6]</td>
<td>4.83</td>
<td></td>
</tr>
<tr>
<td>Planum temporale</td>
<td>[57, -15, 6]</td>
<td>4.62</td>
<td></td>
</tr>
<tr>
<td>Superior temporal lobe*</td>
<td>[-57, -24, 3]</td>
<td>4.53</td>
<td>267</td>
</tr>
<tr>
<td>Planum polare</td>
<td>[-51, -3, 0]</td>
<td>3.77</td>
<td></td>
</tr>
<tr>
<td><strong>Cortical sensorimotor, frontal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary somatosensory cortex, BA3a</td>
<td>[42, -6, 30]</td>
<td>3.58</td>
<td></td>
</tr>
<tr>
<td>Superior frontal gyrus, BA6</td>
<td>[-27, -6, 60]</td>
<td>3.82</td>
<td></td>
</tr>
<tr>
<td>Superior frontal gyrus, BA6</td>
<td>[-9, 6, 69]</td>
<td>3.49</td>
<td></td>
</tr>
<tr>
<td>Middle frontal gyrus</td>
<td>[-51, 9, 48]</td>
<td>3.48</td>
<td></td>
</tr>
<tr>
<td>Premotor cortex, BA6*</td>
<td>[9, 0, 54]</td>
<td>3.97</td>
<td>188</td>
</tr>
<tr>
<td>Precentral gyrus, BA6*</td>
<td>[51, -3, 48]</td>
<td>4.97</td>
<td>136</td>
</tr>
<tr>
<td>Precentral gyrus BA4a*</td>
<td>[-42, -12, 51]</td>
<td>4.32</td>
<td>268</td>
</tr>
<tr>
<td>Anterior insular cortex</td>
<td>[-36, 12, -12]</td>
<td>3.84</td>
<td></td>
</tr>
<tr>
<td><strong>Cortical visual, parietal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calcarine sulcus</td>
<td>[21, -57, 21]</td>
<td>3.37</td>
<td>9</td>
</tr>
<tr>
<td>Precuneus</td>
<td>[-15, -63, 51]</td>
<td>3.74</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>[15, -60, 42]</td>
<td>4.06</td>
<td>46</td>
</tr>
<tr>
<td>Posterior cingulate gyrus</td>
<td>[-3, -33, 45]</td>
<td>3.46</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>[-9, -27, 39]</td>
<td>3.21</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Spatial coordinates of the local hemodynamic activity maxima during echolocation versus null (click production without auditory feedback) in an echolocation expert. MNI-coordinates (p < 0.05 FDR-corrected, spatial extent threshold of 5 voxels), together with the z-score, and the cluster extent are shown (see also Fig. 9B). * Significant after clusterwise FWE-correction (p < 0.05)

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


A  Active vs. Passive Echolocation

Room size ratings

<table>
<thead>
<tr>
<th>IR compression factors</th>
<th>Null</th>
<th>0.2</th>
<th>0.5</th>
<th>0.7</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>Passive</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

B

x = -52  x = 6

x = 52  z = 6
Parametric fMRI signal change in a single voxel

A

B

C

fMRI signal (arb. units)

Room size rating

BRIR compression factor

BRIR amplitude
Echolocation expert

A

B

BRIR c.f.
- 0.2
- 0.5
- 0.7
- 1.0

x = -52
x = 6
z = 6

x = 52