Auditory cue combination in perceptual externalization of static sound sources: a model-based meta analysis

Robert Baumgartner[1] and Piotr Majdak

Acoustics Research Institute, Austrian Academy of Sciences

Abstract

Under natural conditions, listeners perceptually attribute sounds to external objects in their environment. This core function of perceptual inference is often distorted when sounds are produced via hearing devices such as headphones or hearing aids, resulting in sources being perceived unrealistically close or even inside the head. Psychoacoustic studies suggest a mixed role of various monaural and interaural cues contributing to the externalization process. We developed a model framework for perceptual externalization able to probe the contribution of cue-specific expectation errors and to contrast dynamic versus static strategies for combining those errors within static listening environments. The model was applied to various acoustic distortions with constant reverberation as tested in five previous experiments. Our results suggest that the decisions follow a static, weighted accumulation of expectation errors for both monaural and interaural spectral shapes, without a significant contribution of other auditory cues.

Keywords: computational modelling, distal attribution, perceptual decision making, predictive processing, sound externalization, spatial hearing

1 Corresponding author: robert.baumgartner@oeaw.ac.at

Introduction

For a successful interaction with the environment, recent and influential theories suggest that the brain's primary objective is to infer the causes of its sensory input by creating an internal model of the environment generating expectations about the incoming cues [1]. These theories convince by examples coming from various areas of sensory perception and higher-order cognitive functions [2]. The underlying perceptual decision making is usually based on multiple, simultaneously accessible cues. Various decision strategies such as weighted sums or cue selection have been put forward in the context of visual search tasks [3] but remained an unresolved topic in spatial hearing [4]. The goal of this study is to test those concepts on a particularly puzzling perceptual phenomenon of spatial hearing, namely, the collapse of perceptual externalization (or distal attribution) [5], which denotes the lack to associate sensations with external objects [6].

Auditory externalization constitutes a critical aspect of spatial hearing because it can be easily disrupted when listening to sounds, e.g., via headphones or other hearing devices, that do not accurately represent the spatial properties of the listener's natural acoustic exposure [7–9]. Given the high prevalence of daily hearing device usage and the rapid technological advance in augmented reality systems, this topic is timely and of high social relevance.

The spatial properties of the sound arriving at the two ears are manifold and so are the cues that can be used for spatial inference [10]. Many studies aimed to identify in particular spectral cues affecting the perceptual externalization of sound sources [7]. These cues can be categorized as either monaural or interaural. Monaural cues include the overall sound intensity, known as a dominant relative distance cue [11], or the spectral shape, known to be crucial for sound localisation beyond the horizontal plane [12–14]. Interaural cues include the interaural intensity difference (IID), interaural time difference (ITD) and/or interaural coherence (IC), which are well known to be crucial for sound localisation within the left-right dimension [15, 16] but may also affect spatial hearing in other dimensions [17, 18]. Moreover, cues from both categories may potentially be evaluated on a frequency-specific or broadband manner. Besides the multitude of potential cues, a general problem is also that many of those cues co-vary and that it has never been investigated which perceptual decision strategy may underlie the listeners' response behaviour in those psychoacoustic tasks.

While the cue-based encoding can be considered as trivial, probably already happening before reaching the auditory cortex [19], more complex structures are required to combine the cues to form a decision stage yielding the final percept of externalization. Cortical representations contain an integrated code of sound source location [20–22], but also retain independent information about spatial cues [23–26], allowing the system to decide how likely they would have arisen from one or multiple sources [26], or would have arisen from an external source, i.e., manifesting as a final percept of auditory externalization. The decision strategy

used by the auditory system to form externalization is unclear yet. Here, we tested two general types of decision strategies.

The first potential strategy follows the idea that based on exposure statistics our brain has established a rather fixed weighting of information from certain spatial auditory cues in order to perceive an externalized auditory event – independent of the listening context. This can be represented as a static weighted-sum model (WSM, see Fig. 1d) that scales the expectation errors obtained for a single cue with weights adjusted based on long-term listening experience. The WSM is often used to merge neural processing of multiple cues, with examples ranging from a lower, peripheral level such as neural binaural processing [27] over higher cortical levels integrating visual orientation cues [28] to even more abstract levels such as the representation of person identities [29]. Once the weights are settled, the WSM has the advantage of being a static process with no need for further adaptation during the externalization process.

The second potential decision strategy is of selective nature and has been promoted, for instance, in the context of visual search [30, 31] or audio-visual dominance [32]. The idea is that, depending on the incoming stimulus, our brain selects the one of the externalization-related cues that fulfils or breaks one or most of the listener's prior expectations (see Fig. 1e). We considered three variants of such a selection strategy. First, a minimalist approach would assume that a sound can be externalized if at least one cue promotes externalized perception by matching the listener's expectations. We implemented this as a winner-takes-all (WTA) strategy that considers the largest externalization rating (minimum expectation error across all individual cues) as the total externalization rating. Second, the contrary is the perfectionist approach, which requires all of the cues to promote an externalized percept – the cue breaking its expectation the most will dominate the perception. This is implemented as the loser-takes-all (LTA) strategy, which considers the smallest externalization rating (maximum expectation error across cues) as the total rating. Third, the conformist approach is intermediate to these two extremes and selects the cue-specific expectation error being most consistent to all others. We implemented this as a median-takes-all (MTA) strategy, which considers the median across all single-cue-based ratings as the total externalization rating.

Here, we propose a model framework allowing us to disentangle those decision strategies potentially involved in the multi-cue-based task of auditory externalization. We used that framework to simulate five experiments from representative psychoacoustic externalization studies [17, 33–35], all focusing on the effects of spectral distortions under spatially static listening conditions, but differing considerably in manipulation method and bandwidth. The model represents each listener's expectations for perceptual inference as internal templates and uses them for comparison with incoming cue-specific information. The decision strategies were not only compared for the whole set of cues, but also for a successively reduced set, in order to address the potential redundancy between them. Our results suggest

that auditory externalization judgments under spatially static situations with constant reverberation underlie a fixed weighting scheme between monaural and interaural spectral cues, rather than a dynamic selection of those auditory cues.

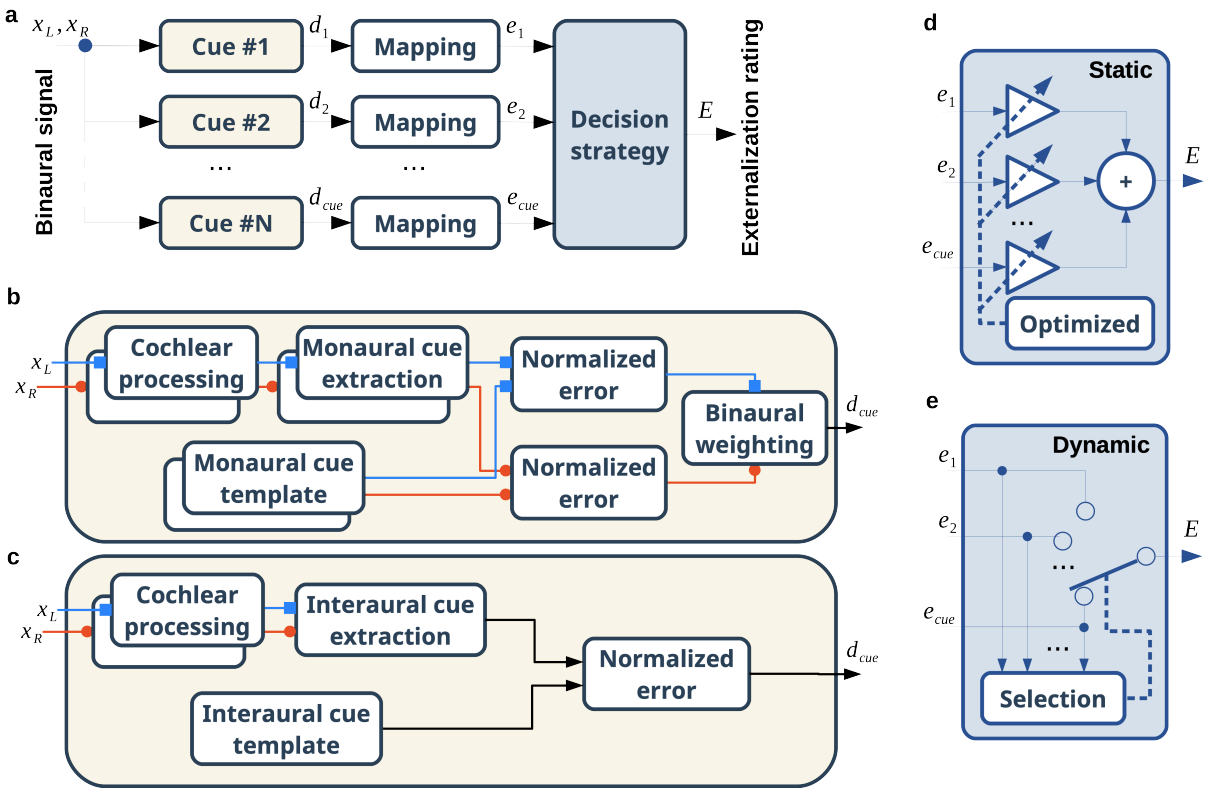

***Fig. 1 | General structure of the sound externalization framework. a**, Processing of the binaural signal x_L, x_R resulting in a cue-based error d_{cue}, a stage mapping the cue-based error to a normalized externalization rating e_{cue}, and a decision stage combining the cue-specific ratings to a final externalization rating E. **b**, Processing of the binaural signal to calculate a single-cue error based on monaural cue templates. **c**, Processing based on interaural cue templates. **d**, Decision stage based on a static weighted-sum model (WSM). **e**, Dynamic decision stage based on the selection of a cue-specific rating. Minimalist (LTA), conformist (MTA), and perfectionist (WTA) approaches were considered for selection (see text for further explanation).*

Materials and Methods

Structure of the model mapping signals to externalization ratings

The model flexibly comprises a variety of cues as summarized in Tab. 1. Depending on the considered cue, slightly different model architectures were required for processing either monaural (Fig. 1b) or interaural cues (Fig. 1c). Both architectures follow a standard template matching procedure [7, 36–38] as they take a binaural signal as an input and simulate externalization ratings as an output after performing comparisons with internal cue templates. Those templates were derived from listener-specific head-related transfer functions (HRTFs) or binaural room impulse responses (BRIRs), depending on the modelled experiment.

Property	Monaural (M_)	Interaural (I_)
Spectral Shape (_SS)	MSS	ISS
Spectral Standard Deviation (_SSD)	MSSD	ISSD
Coherence (_C)	-	IC
Time-Intensity Trading (_TIT)	-	ITIT
Intensity (_I)	MI	-

Tab. 1 | Potential auditory externalization cues considered for the evaluation.

For both monaural and interaural cues, the template comparison may either be based on whole spectral shapes or scalar spectral metrics. Hence, in addition to monaural spectral shape (MSS) [12, 39] and interaural spectral shape (ISS) [17], we considered various cues that could be associated with externalization or distance perception of anechoic sounds: the difference in broadband monaural intensity (MI), the monaural spectral standard deviation (MSSD) [34, 40], the spectral standard deviation of IIDs (ISSD) [41], the broadband interaural coherence (IC) [42], and the inconsistency between ITD and IID (ITIT) [18].

The model framework (Fig. 1a) consists of 1) cue-processing stages, each of them encoding the incoming binaural signal, $[x_L, x_R]$, into a single-cue expectation error, d_{cue}, 2) stages mapping that single-cue error to a normalized externalization rating, e_{cue}, and 3) a decision stage combining the single-cue based ratings into the final externalization rating, E. The mapping from d_{cue} to e_{cue} was scaled individually for each experiment, unless otherwise noted.

Error metrics

For the MI cue, the overall level differences were considered as:

$$d_{MI} = \Delta MI / MI_{template} \,, \tag{1}$$

with $\Delta MI = |MI_{target} - MI_{template}|$ denoting the difference in root mean square (RMS) levels (in dB) of the incoming broadband signals. Differences smaller than ±1 dB were considered to be below the just-noticeable difference [43] and thus set to zero. The error based on MI, d_{MI} was calculated for each ear separately and then averaged across both ears.

The IC cue was calculated as $IC = \lim_{\tau} \inf \int x_L(t-\tau) x_R(t)\, dt$ within the range of $\tau \in [-1,1]$ ms. The error based on IC was then calculated by comparing the target IC with the template IC:

$$d_{IC} = \frac{|IC_{target} - IC_{template}|}{IC_{template}} \,. \tag{2}$$

The errors for the other cues were calculated after filtering the target and template signals through a bank of fourth-order Gammatone filters with a regular spacing of one equivalent rectangular bandwidth [44]. The spectral excitation profiles were computed from the logarithm of the RMS energy within every frequency band [12, 45]. Audibility thresholds for band-limited signals were roughly approximated by assuming a sound pressure level of 70 dB and a within-band threshold of 20 dB. Further assuming stationary input signals, the spectral profiles were averaged over time, yielding spectral profiles as a function of frequency band, $p(f)$. We chose this efficient approximation of auditory spectral profile evaluation as we observed in previous model investigations that a physiologically more accurate but less efficient approximation of the auditory periphery [46] lead to similar profiles and equivalent predictions of sound localisation performance [39].

The interaural difference of the spectral profiles yielded IIDs as a function of frequency band, $IID(f)$. For the ISSD cue, the model evaluated standard deviations (SD) of $IID(f)$ and computed the negative difference of these deviations between the target and template, relative to the template deviation:

$$d_{ISSD} = \frac{\left| SD_f(IID_{target}(f)) - SD_f(IID_{template}(f)) \right|}{SD_f(IID_{template}(f))} . \tag{3}$$

For the ISS cue, the absolute values of frequency-specific differences between the target and template IIDs were evaluated. Then differences smaller than 1 dB were set to zero and larger differences, $\Delta|IID(f)|$, were normalized by the template IIDs and averaged across frequency bands, yielding:

$$d_{ISS} = \frac{1}{N_f} \sum_f \frac{\Delta|IID(f)|}{|IID_{template}(f)|} , \tag{4}$$

with N_f being the number of frequency bands.

The ITIT expectation error was calculated as the broadband deviation between target-to-template ratios of ITD and IID:

$$d_{ITIT} = \left| \frac{\Delta ITD}{ITD_{template}} - \frac{\Delta IID}{IID_{template}} \right| , \tag{5}$$

with $\Delta ITD = ITD_{target} - ITD_{template}$ and $\Delta IID = IID_{target} - IID_{template}$. ΔITD and ΔIID smaller than ±20 µs and ±1 dB, respectively, were set to zero. The ITDs were derived from binaural signals following a procedure, in which the signals were low-pass filtered at 3 kHz and the ITD was the time lag that yielded maximum IC of the temporal energy envelope [47].

For the MSS and MSSD cues, positive spectral gradient profiles were derived exactly as in our previous work [39]. Briefly, first monaural spectral gradients were obtained by differentiating the excitation profiles ($p(f) \rightarrow p'(f)$) and softly restricting the value range by an elevated arctangent:

$$MSG(f) = \arctan(p'(f) - \pi/2) + \pi/2 \,. \tag{6}$$

For the MSS cue, these gradients were then compared between the target and template separately for each ear by applying the same procedure as for the ISS metric, that is, calculating absolute target-to-template differences, normalizing differences larger than 1 dB $(\Delta|MSG(f)|)$ by the template gradients, and averaging those differences across frequencies:

$$d_{MSS,L/R} = \frac{1}{N_f} \sum_f \frac{\Delta|MSG_{L/R}(f)|}{|MSG_{template,L/R}(f)|} \,, \tag{7}$$

separately for the left and right ear as indexed by L and R, respectively. The MSS error metric was defined in analogy to ISSD:

$$d_{MSSD,L/R} = \frac{|SD_f(MSG_{target,L/R}(f)) - SD_f(MSG_{template,L/R}(f))|}{SD_f(MSG_{template,L/R}(f))} \,. \tag{8}$$

These unilateral error metrics were then combined according to a binaural weighting function [12, 13], effectively increasing the perceptual weight of the ipsilateral ear with increasing lateral eccentricity:

$$d = d_R + \frac{d_L - d_R}{1 + e^{-\phi/\Phi}} \,, \tag{9}$$

with $\phi \in [-90°, 90°]$ denoting the lateral angle (left is positive) and $\Phi = 13°$.

Mapping to externalization ratings

A sigmoidal mapping function scaled by $2e_{range}$, shifted by e_{offset}, and slope-controlled by a sensitivity parameter S_{cue} was used to map the metrics of expectation error d_{cue} to externalization ratings e_{cue}:

$$e_{cue} = \frac{2e_{range}}{1 + \exp(d_{cue}/S_{cue})} + e_{offset} \,. \tag{10}$$

The nominator was doubled because the rating scale used in previous experiments was usually one-sided with respect to the reference sound, i.e., listeners were asked to rate only decreases and not increases of perceived externalization.

The mapping function in Eq. (10) contains one free model parameter, S_{cue}, inversely related to the slope of the function mapping changes in the deviation metrics to changes in the externalization ratings. The mapping sensitivity is denoted by $1/S_{cue}$ because larger $1/S_{cue}$ yield steeper mapping functions that project smaller errors to smaller externalization ratings. This cue- and experiment-specific sensitivity was obtained by minimizing the squared simulation error, which was defined as the RMS of differences between actual and simulated externalization ratings (normalized scales). For the minimization, we applied the Nelder-Mead simplex (direct search) method (fminsearch, Matlab Optimization Toolbox, The Mathworks Inc.).

Decision stage

For the WSM, optimal weights for scaling the contribution of a cue to the final rating were obtained by minimizing the simulation error. We used the same optimization technique as used for S_{cue}. Weights smaller than 0.005 were considered as negligible and were set to zero.

For a fair comparison across our simulations of the decision strategies, the same number of model parameters was considered in the optimization. For the dynamic strategies, the mapping parameters were optimized to best fit the data of all experiments. For the weighted-sum strategy, the mapping parameters were fixed and corresponded to those from single-cue simulations, and the individual summation weights were optimized to best fit the data of all experiments.

Considered studies

We simulated results of five previous headphone experiments [17, 33–35]. The pool of experiments was a compromise of data availability and the degree to which the test conditions isolate specific cues.

In Exp. I and II, Hartmann and Wittenberg [33] synthesized the vowel /a/ with a tone complex consisting of 38 harmonics of a fundamental frequency of 125 Hz, yielding a sound limited up to 4750 Hz. This sound lasted for 1 s and was presented via headphones and filtered with individualized HRTFs corresponding to 37° azimuth (0° elevation). The HRTFs were manipulated at the frequencies of individual harmonics and the listeners rated the perceived externalization of those sounds presented via headphones on a labeled four-point scale (ranging from “0. The source is in my head” to “3. The source is externalized, compact, and located in the right direction and at the right distance”). The loudspeaker used to measure each listener’s HRTFs remained visible throughout the experiment. For modelling the average results of these experiments, we used HRTFs from 21 exemplary listeners contained in the ARI database.

In Exp. III, Hassager et al. [17] investigated the effect of spectral smoothing on the auditory externalization of Gaussian white noises, band-limited from 50 Hz to 6000 Hz and lasting about 4 s. These sounds were filtered with individualized BRIRs (reverberation time of ~0.2 s) in order to simulate sound sources positioned at azimuths of 0° and 50°. As independent experimental variable, Gammatone filters with various equivalent rectangular bandwidths (ERBs) were used to spectrally smoothen either the direct path portion (until 3.8 ms) or the reverberation of the BRIRs. Filters with larger ERBs more strongly smoothed the shape of the magnitude spectrum and reduced the degree of externalization only when applied to the direct path. Smoothing applied to the reverberation was perceptually ineffective. Similar to Exp. I and II, listeners responded on a five-point scale where they perceive the sound as coming from (inside the head denoted as 1 and the visible loudspeaker position denoted as 5). Because the original BRIRs were not accessible, our model simulations were

again based on the same 21 (anechoic) HRTFs from the ARI database and only addressed the effect of spectrally smoothing the direct path.

In Exp. IV, spectral smoothing was applied to Gaussian white noise bursts band-limited to higher frequencies from 1 kHz to 16 kHz [34]. Within that frequency range, the listeners' HRTFs used for binaural rendering were applied as individually measured (C = 1), as being spectrally flat (C = 0), or as something in between with reduced spectral contrast (C = 0.5). Pairs of about 0.5-s-long sounds were presented and the listeners were asked to judge whether the second sound was perceived closer or farther than the first sound. We evaluated only data from their "discontinuous trial" condition (of their Exp. II) with an inter-stimulus interval of 100 ms, which did not elicit an auditory looming bias and thus allowed to estimate absolute externalization ratings from these paired binary judgments by calculating mean relative frequencies of "farther" judgments. Two out of twelve listeners that unexpectedly perceived the spectrally flat sound (C = 0) as being farther than the individualized reference sound (C = 1) were not included in the present model analysis.

In Exp. V, Boyd et al. [35] used individualized BRIRs to simulate a talker positioned at 30° azimuth. The speech samples consisted of short sentences lasting about 3 seconds and provided sufficient speech energy up to 15 kHz. The study compared externalization ratings for in-the-ear (ITE) vs. behind-the-ear (BTE) microphone casings as well as broadband (BB) vs. 6.5-kHz-low-pass (LP) filtered stimuli at various mixing ratios with stereophonic recordings providing only an ITD but no IID. The listeners were asked to rate the degree of externalization against a preceding reference sound (1 s inter-stimulus interval) on a continuous scale with five labels (ranging from "0. Center of head" to "100. At the loudspeaker"). The reverberation time of 0.35 s remained constant across mixing ratios. For our model simulations, original BRIRs were only available for three out of seven (normal-hearing) listeners.

As summarized in Tab. 2, these five experiments differed in whether they tested in anechoic or reverberant environments and provided visual information about the reference source or not. While the constant reverberation and visual information may have stabilized auditory externalization, the original results show that they did not prevent the tested signal modifications to be effective within the tested condition. Hence, we here assumed that such

Exp	Signal	Bandwidth	Duration	Azimuth	Reverb	Visible	Data
I&II	Tone complex	0.1 – 5 kHz	1 s	37°	none	yes	Avg.
III	White noise	0.05 - 6 kHz	4 s	0°, 50°	200 ms	yes	Avg.
IV	White noise	1 - 16 kHz	0.5 s	0°, ±90°	none	no	N = 10
V	Speech	varied	3 s	30°	350 ms	yes	N = 3

Tab. 2 | Differences between considered studies with respect to methodological details and data availability. *Visible refers to the visibility of a loudspeaker at the intended source location. Model simulations were either compared with the reported average results retrieved from the original publications (Avg.) or on the basis of individual results from N listeners.*

methodological differences do not affect our effects of interest in a substantial way that cannot be compensated by an individually adjusted mapping parameter (S_{cue}).

For ease of comparability, we normalized the response scale of every individual experiment to externalization rating scores ranging from zero to one. These scores should be interpreted as the degree of externalization, mediating the egocentric distance percept [7], and not as the level of perceptual ambiguity per se.

Results

Individual cues: large differences across experimental conditions

We first investigated the ability of each individual cue to explain the effects of signal manipulations on auditory externalization as tested in the five experiments [17, 33–35]. Figure 2 shows the simulated externalization ratings together with the actual ratings from the original experiments. The frequency ranges of tested stimuli were classified as high or low with respect to the lower frequency bound around 5 kHz at which spectral cues can be induced by normal pinna morphologies.

Exp. I: Effect of IID modifications at low frequencies

In Exp. I, the tone magnitudes up to a certain harmonic of a complex tone were set to the interaural average, effectively removing IIDs up to that harmonic's frequency (Fig. 2a, Exp. I) [33]. When simulating this experiment, the monaural spectral cues (MSS and MSSD) showed a non-monotonic relationship across the modified frequency ranges. Those simulated ratings largely diverged from the actual ratings especially for the largest frequency ranges and resulted in large simulation errors (Fig. 2e, Exp. I). Further inspection showed that the modification of interaural averaging only marginally affected those monaural cues because at low frequencies the complex acoustic filtering of the pinnae is negligible and thus monaural spectral shapes are quite similar at both ears. In contrast, the broadband monaural cue (MI) was able to explain the actual ratings surprisingly well because the IID removal caused a small but systematic increase in sound intensity, being in line with the systematic decrease in externalization ratings.

Most interaural cues (ISS, ITIT, and IC) were able to explain the actual ratings very well. Differences in one interaural cue (IC) were, however, very small and thus required a very steep mapping function in order to become informative (indicated by a low sensitivity in Fig. 2e, Exp. I). Simulations based on interaural spectral contrasts (ISSD) failed presumably because the evaluation of the standard deviation metric became unstable for spectral IID distributions strongly deviating from a normal distribution. Overall, the broadband interaural cue (ITIT) performed best with the minimum RMS simulation error of 0.06.

Exp. II: Effect of monaural modifications at low frequencies

In Exp. II, instead of removing interaural differences, the ipsilateral magnitude spectrum of the complex tone was flattened up to a certain harmonic while the contralateral magnitudes were shifted in parallel [33]. This procedure effectively maintained the original IIDs but changed the monaural spectral profiles and by doing so also gradually distorted perceived externalization (Fig. 2a, Exp. II).

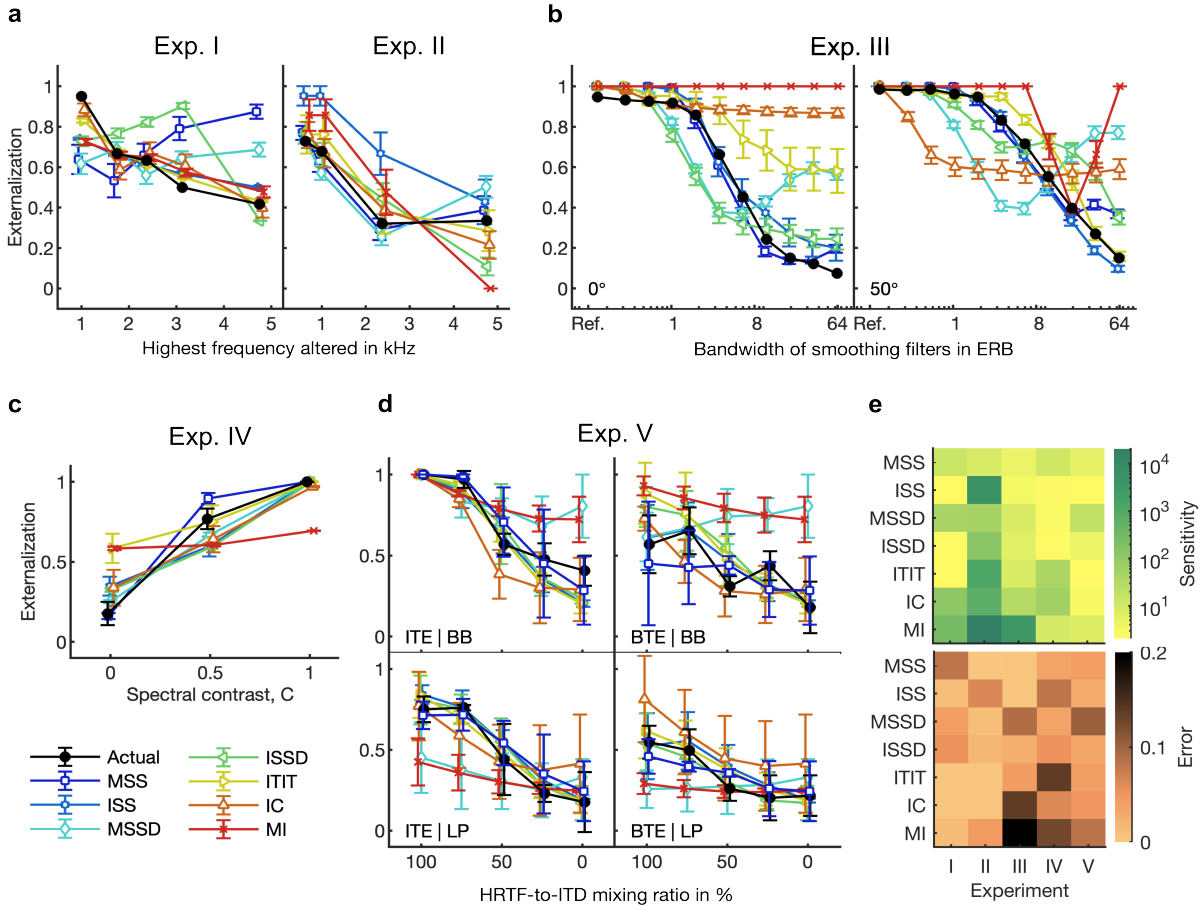

Fig. 2 | Externalization ratings: actual data from psychoacoustic experiments *(closed circles)* ***and simulations of the single-cue models*** *(open symbols).* ***a****, Effects of low-frequency modifications tested by Hartmann and Wittenberg [33]. Exp. I: IID set to zero (bilateral average magnitude); actual data from their Fig. 7, average from N = 2. Exp. II: ipsilateral magnitudes flattened (IID compensated by contralateral magnitudes); actual data from their Fig. 8, average from N = 4. Simulated results for various cues, average from N = 21.* ***b****, Exp. III: effect of spectral smoothing of low-frequency sounds presented from various azimuths (left: 0°; right: 50°); actual data represents direct-sound condition from Hassager et al. [17], average from N = 7. Simulated N = 21.* ***c****, Exp. IV: effect of spectral smoothing in high frequencies as a function of spectral contrast (C=1: natural listener-specific spectral profile; C=0: flat spectra); actual data calculated from the paired-comparison data from Baumgartner et al. [34], N = 10 (actual and simulated).* ***d****, Exp. V: effects of stimulus bandwidth and microphone casing for various mixes between simulations based on listener-specific BRIRs (100%) and time-delay stereophony (0%); actual data from Boyd et al. [35], N = 3 (actual and simulated). ITE: in-the-ear casing; BTE: behind-the-ear casing; BB: broadband stimulus; LP: low-pass filtered at 6.5 kHz; Error bars denote standard errors of the mean.* ***e****, Cue-specific sensitivities and RMS simulation errors revealed as best explaining the actual externalization ratings. Per experiment, the smallest error indicates the most informative cue.*

In contrast to Exp. I, the simulations based on monaural spectral cues (MSS and MSSD) reflected the decrease in actual ratings very well, as indicated by RMS simulation errors as low as 0.05 (for MSS). As expected, simulations based on interaural cues failed because the degradation induced by flattening the ipsilateral spectrum was designed to maintain the original interaural features.

Exp. III: Effect of spectral smoothing at low frequencies

The spectral smoothing applied in Exp. III affected the auditory externalization of noise more gradually for the lateral as compared to the frontal direction (Fig. 2b) [17]. Again, only some of the cues were able to explain the systematic trend of externalization degradation. The stimulus manipulation affected some cues only hardly (ITIT, IC, and MI) or in a non-systematic manner (MSSD) whereas both monaural and interaural spectral shape cues (MSS and ISS) yielded more accurate simulations (RMS errors of 0.18 and 0.25, respectively). Both cues yielded simulations consistent with the actual results in that they were insensitive to spectral smoothing below one ERB. It appears noteworthy that the monaural (MSS) outperformed the interaural (ISS) cue in this particular experiment, which has been used to promote an earlier modelling approach only based on the interaural cue (ISS) [17]. Moreover, neither cue was able to explain why actual ratings were slightly below the maximum for the frontal reference sound (and bandwidths ≤ 1 ERB). Additional factors not represented in the model seem to be necessary to explain this lateral dependence of reference externalization.

Exp. IV: Effect of spectral smoothing at high frequencies

Focusing on the high-frequency range from 1 kHz to 16 kHz, where the pinnae induce the most significant directional spectral variations, Exp. IV concordantly showed that spectral smoothing degrades externalization perception (Fig. 2c) [34]. Broadband cues (ITIT and MI) were hardly affected by this method and were not able to explain the actual results. Best simulations were obtained with monaural spectral cues (MSS and MSSD), followed by the interaural cues (ISS, ISSD, and IC). Hence, the template comparison based on monaural spectral shapes (MSS) allowed accurate simulations of the effects of spectral smoothing independent of both the particular smoothing methods and frequency ranges considered in Exps. III and IV.

Exp. V: Bandwidth limitation and sensor displacement

Exp. V presented broadband speech samples from a mid-left position and compared externalization ratings for different sensor positions, stimulus bandwidths, and mixing ratios that eliminated IID. The simulated ratings showed much variability across cues and conditions (Fig. 2d). The broadband BTE condition caused the most distinct spectral modifications and was particularly informative about the explanatory power of considered cues. Most cues were able to follow the actual ratings quite well (except MSSD and MI). On average across condi-

tions, expectations based on interaural spectral templates (ISS, simulation error of 0.16) seem to have been most relevant to the listeners in this experiment.

Individual cue summary

The overall picture of simulation errors suggests that there is no single cue able to explain the externalization ratings for all experiments but that there is a small set of cues which listeners seem to base their externalization ratings on. These cues correspond to the evaluation of monaural and interaural spectral features (MSS and ISS, respectively) as well as the broadband interaural disparities between ITD and IID (ITIT), respectively. This is consistent with previous observations showing that both interaural and monaural changes affect externalization perception.

The monaural cue (MSS) yielded best simulations in three out of five experiments as well as on average across the five experiments (simulation error of 0.18). The evaluation of this cue focuses mainly on positive gradients in the spectral profile as motivated by physiological findings in the cat dorsal cochlear nucleus [48, 49]. To assure that our results were not confounded by a suboptimal choice of gradient sign, we repeated all simulations also with two different model configurations either favouring negative or considering both positive and negative gradients (technically, we implemented this by either switching the signs or removing the $\pm\pi/2$ shifts in Eq. (6), respectively). We found that the average simulation error increased to 0.24 (for negative gradients only) and to 0.23 (for both negative and positive gradients), consolidating our choice of focusing the evaluation on positive spectral gradients.

The sensitivity parameters used to scale the mapping function from cue-specific error metrics to externalization scores were optimized separately for every cue and experiment. The separate optimization was reasoned by the limited quantitative comparability of subjective externalization ratings because of various factors such as differently trained subjects, different contexts, different experimental procedures (especially with respect to presenting reference stimuli), and other methodological differences. Nevertheless, the optimization procedure yielded somewhat similar sensitivity parameters for the same cue across experiments whenever that cue was informative as indicated by small RMS simulation errors.

In summary, we found considerable variance in cue-specific simulation accuracy. Simulations based on a single monaural cue (MSS) turned out to be most potent in explaining the set of the considered experiments. However, the best cue clearly depended on the experiment and future simulations based on that particular cue would fail in situations similar to Exp. I. This indicates that the auditory system does not simply rely on a single cue and leads to our second research question of which perceptual decision strategy is used to combine cues in different listening situations.

Decision strategy: static weighting outperforms dynamic selection

We aimed to determine which decision strategy best explains externalization ratings across all experiments without an a-priori knowledge about the context. In the static decision strategy (WSM), the simulated listener derives the final externalization rating from a linearly weighted combination of single-cue ratings (Fig. 1b). The weights were obtained from an optimization procedure minimizing the simulation errors. The dynamic decision strategies, that is, WTA, MTA, and LTA, selected the largest, median, and smallest simulated externalization rating, respectively, obtained across the considered cues. By doing so, the cue weights become binary: the cue providing the largest, median, or smallest externalization rating is considered as the total rating, all other cues are ignored.

Overall, the static cue combination (WSM) outperformed the dynamic strategies, as indicated by significantly lower simulation errors (Fig. 3a). It simulated the externalization ratings based on the weighted sum of mainly two spectral cues: a monaural cue (MSS) weighted by about two thirds and an interaural cue (ISS) weighted by the remaining third, while the other cues contributed only marginally (Fig. 3b). The contribution of the interaural cue is in line with previous findings in the context of Exp. III, for which only interaural cues have previously been used to explain externalization ratings [17]. Across other experiments, as shown here, the monaural cue becomes essential.

Within the dynamic strategies, the conformist approach (MTA) performed significantly better than all the other selection approaches (Fig. 3a). The minimalist approach (WTA) performed particularly poorly, being close to the chance rate of 0.5. The cues selected by the

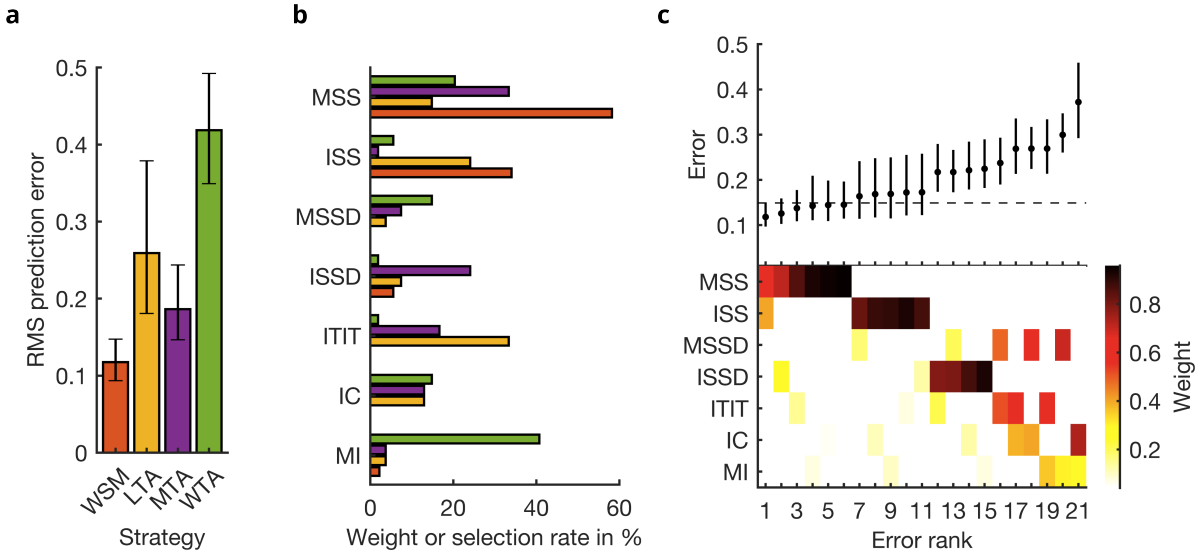

Fig. 3 | Simulation errors for different decision strategies and cue combinations show that static combination (WSM) based on monaural and interaural spectral shape cues (MSS, ISS) performs best. a, *Simulation errors for different strategies and pooled experimental data, N = 54. Error bars denote 95% confidence intervals estimated via bootstrapping (1000 resamples). WSM: weighted-sum model; L/M/WTA: looser/median/winner takes all.* ***b,*** *Individual cue contributions. Cue abbreviations defined in Tab. 1.* ***c, Top:*** *Simulation errors for pairwise combinations of considered cues. Dashed line shows the border of significant difference to the best pair (MSS and ISS).* ***Bottom****: Considered cue pairs with their respective weights.*

dynamic strategies contributed more diversely than the weights of the static strategy (WSM, Fig. 3b). In the MTA strategy, the monaural cue (MSS), also favoured by the static strategy, played an important role, suggesting that it provides the most consistent ratings across all considered cues. In the LTA strategy, broadband interaural cues (ITIT) were most frequently selected, which is in accordance with our single-cue simulation results where this particular cue also provided good results for some conditions. Concordantly, the large variance in simulation errors (large error bars for LTA in Fig. 3a) indicates here that such a break down of broadband interaural disparities is only meaningful in some listening conditions.

Cue relevance: both monaural and interaural spectral shapes matter

Our analysis suggests that sound externalization is mainly based on two cues. However, most of the cues co-vary to some degree, and thus the relevance of a particular cue as a contributor to the cue combination may be strongly affected by the joint consideration of dependent cues. To investigate the effect of such interdependency, we simulated listeners using the static decision strategy (WSM) with only two cues but all possible combinations.

The results, sorted by increasing simulation error, are shown in Fig. 3c. As expected there is a considerable variance in the simulation errors obtained by different pairs, with an order of magnitude between the best and worst pair. The condition considering the previously favoured cues (MSS and ISS) yielded the smallest simulation error, confirming our findings from simulations considering more than two cues.

The results underline the importance of both cues by two means. First, all pairs including the monaural cue (MSS) were ranked highest and the simulation error remained small regardless of the other considered cue. As long as that cue was involved, the simulation errors did not increase significantly (remained within the confidence intervals of the best condition). Thus, the MSS cue seems to be the most important cue for auditory externalization under static listening conditions with fixed reverberation. Second, by not considering that monaural cue, the error increased (with the mean above the confidence intervals of the best condition), but was still not significantly different to the best condition, as long as the spectral interaural cue (ISS) was considered in the modelling. In fact, the simulation errors significantly increased ($p < 0.05$, compared to the best condition) as soon as neither the monaural (MSS) nor the interaural (ISS) cues of spectral shape were considered. Other spectral cues that only evaluate the amount of spectral contrast (MSSD and ISSD) instead of its shape fail to explain perceived externalization.

Discussion

In order to uncover the essential cues and the decision strategy underlying auditory externalization perception in various static listening situations, we developed a template-based modelling framework and applied it to a set of psychoacoustic experiments investigating how

spectral distortions degrade the auditory externalization percept. Our results suggest that a static, weighted combination (WSM) rather than a dynamic selection (LTA, MTA, WTA) of monaural and interaural spectral cues drives perceptual decisions on the degree of auditory externalization. Hence, although listeners are sensitive to many individual cues, their externalization perception seems to be driven by a static combination of only few cues with fixed perceptual weights.

Dominant cues

The two major cues favoured by our model selection procedure are both based on spectral shapes. Monaural (MSS) [50] as well as interaural (ISS) [51] spectral cues are known to be informative about absolute distance in the acoustic near-field of a listener. Those two cues are also well-known to be important for and complementary in the process of sound localisation. Monaural spectral shapes constitute localisation cues within sagittal planes, i.e., including both the vertical and front/back dimension [38]. The current understanding of this cue is that it resembles the processing of monaural positive spectral gradient profiles (MSS). In line with previous assumptions [12, 39] and electrophysiological measurements in the dorsal cochlear nucleus [48], reducing the contribution of negative spectral gradients here turned out to improve our simulations of perceived externalization.

Interaural intensity differences are very well established as a cue for auditory lateralization [15, 16]. Given the inherent frequency selectivity of the auditory periphery it is reasonable to assume that relative differences to an internal reference are evaluated on a frequency-specific level [17] as implemented in the ISS cue of our model. On the other hand, the accessibility or predictability of such an interaural spectral evaluation must be limited somehow in the auditory system because otherwise it would not make sense to consider the monaural counterpart at all for spatial inference, as this cue intrinsically suffers from ambiguity with the source spectrum [52].

Template matching in view of predictive processing

In order to assess incoming cues most efficiently and rapidly, theories of predictive processing assume a hierarchical propagation of error signals representing differences between internal predictions (templates) and sensory evidence on increasingly complex matters, while errors are weighted by the expected sensory reliability for that matter [53]. In view of auditory externalization, the ultimate goal is to assign the most realistic spatial position to an auditory object and errors between expected and observed cues need to be weighted by the amount of information those cues offer for the inference of an externalized sound-source position.

It was unclear how flexibly those weights may be adjusted. Our results suggest that the weighting between spatial spectral cues, which are naturally persistent, remained fixed across the various experimental contexts. In contrast, monaural intensities, for instance, inform distance inference only on a relative scale with intrinsic ambiguity. The weighting of

such naturally less persistent cues was outside the scope of this study but would be expected to very much depend on the listening context.

Limitations and directions for future investigations

Regarding less persistent cues, reverberation is usually omnipresent but also subject to frequent changes. Reverberation smears the spectral profile and decorrelates the binaural signal, effectively increasing the variance of interaural cues and affecting the reliability of spectral cues. The degree of externalization increases with the amount of reverberation if the reverberation-related cues are consistent with the listener's expectations about the room acoustics [54]. In parallel to the present work, we investigated also the impact of situational changes of the amount of reverberation on externalization perception by extending the here promoted framework to reverberation-related cues [55].

The externalization models proposed here and there do not explicitly consider interactions with directional sound localisation [56–59]. However, the strong overlap of cues between the perceptual externalization and directional sound localisation suggests a strong overlap of the underlying perceptual processes. In our simulations, template comparisons were only performed for the reference direction, ignoring that there might be a strong match to a template from another direction yielding strong externalization at that other perceived location. For example, spectral cue changes may have elicited changes in the perceived polar angle within a sagittal plane. On the other hand, in most of the considered studies, the experimental paradigms involved rating of externalization against a fixed reference stimulus and, thus, one can assume that listeners were able to ignore directional localisation changes. Nevertheless, joint assessment of externalization and directional localisation in future behavioural experiments is needed to further our understanding of the interdependency between those two perceptual processes.

As for directional localisation, the perceptual inference underlying auditory externalization can also adapt to changes in the naturally very persistent spectral cues [60]. Given a fixed weighting of those cues in the model framework, this adaptivity should be solely grounded on the generation of new expectations, i.e., an updating of templates. How such new templates evolve and compete with existing templates would be an exciting objective for future research.

A more substantial extension of the model framework will be required in order to explain perception in dynamic listening situations. Especially self-generated movements like head rotations are known to elicit strong expectations on its acoustic consequences and drastically degrade externalization if not being fulfilled [56, 58, 61]. Internal generative models are considered to constantly shape the listener's expectations by means of either explicit or implicit predictive coding [53]. Embedding the here proposed model in a larger predictive coding framework and using Bayesian inference to account for sensory uncertainty [62] may pave

the road for future work more deeply investigating the short-term dynamics and multisensory context in spatial hearing.

From a technical perspective, our present model successfully predicts auditory externalization ratings in static listening conditions without prior knowledge about the context – a feature often requested by engineers working on audio rendering and human interfaces within the context of augmented and virtual reality [63–65]. In particular, it can be applied to efficiently assess perceived externalization in a variety of acoustic applications relying on spectral cue modifications such as headphones and hearing aids [7]. To simplify the application and further development of the here proposed model, and guarantee reproducibility of results, implementations of both the model (baumgartner2021) and the model simulations (exp_baumgartner2021) are publicly available as part of the auditory modeling toolbox (AMT, version 1.0).

Conclusions

Perception of externalized events depends on the alignment of observed auditory cues with the internal expectations of the listener. In order to uncover the most essential cues and the strategy used by the listener to integrate them, we compared various model types in their capability to explain five experiments focussing on the effects of spectral distortions within a spatially static environment. Most accurate simulations were obtained for the combination of monaural and interaural spectral cues with a fixed relative weighting (about 60% monaural, 40% interaural). In contrast, the dynamic selection of cues depending on its situational magnitude in expectation error was less successful in explaining the experimental results. Together, this suggests that at least in spatially static environments, listeners jointly evaluate both monaural and interaural spectral cues to perceptually infer the spatial location of the auditory object. Further, the proposed model framework appears useful for predicting externalization perception in binaural rendering applications.

Acknowledgements

We want to thank Bill Withmer for kindly providing parts of the original data from Boyd et al. (2012) as well as Henrik Gert Hassager, Barbara Shinn-Cunningham and H. Steven Colburn for fruitful discussions. This work was supported by the Austrian Science Fund (FWF, grant J 3803-N30 to R.B.), Oculus VR LLC, and the European Commission (grant 691229).

Author Contributions

R.B. conceptualized and implemented the research, analysed the results, and wrote the original draft; R.B. and P.M. interpreted the results and revised the manuscript.

Competing Interests

The authors declare no competing interests.

References

1. Friston, K.: A theory of cortical responses. Philos. Trans. R. Soc. B Biol. Sci. 360, 815–836 (2005). https://doi.org/10.1098/rstb.2005.1622
2. Den Ouden, H.E., Kok, P., De Lange, F.P.: How Prediction Errors Shape Perception, Attention, and Motivation. Front. Psychol. 3, (2012). https://doi.org/10.3389/fpsyg.2012.00548
3. Gardner, J.L.: Optimality and heuristics in perceptual neuroscience. Nat. Neurosci. 22, 514–523 (2019). https://doi.org/10.1038/s41593-019-0340-4
4. van der Heijden, K., Rauschecker, J.P., de Gelder, B., Formisano, E.: Cortical mechanisms of spatial hearing. Nat. Rev. Neurosci. 20, 609–623 (2019). https://doi.org/10.1038/s41583-019-0206-5
5. Loomis, J.M.: Distal attribution and presence. Presence Teleoperators Virtual Environ. 1, 113–119 (1992)
6. Weber, E.H.: On the circumstances under which one is led to refer sensations to external objects. In: Proceedings of the royal saxon society for science in Leipzig. pp. 226–237. , Leipzig, Germany (1848)
7. Best, V., Baumgartner, R., Lavandier, M., Majdak, P., Kopčo, N.: Sound Externalization: A Review of Recent Research. Trends Hear. 24, 2331216520948390 (2020). https://doi.org/10.1177/2331216520948390
8. Blauert, J.: Spatial hearing. The Psychophysics of Human Sound Localization. MIT-Press, Cambridge, MA (1997)
9. Durlach, N.I., Rigopulos, A., Pang, X.D., Woods, W.S., Kulkarni, A., Colburn, H.S., Wenzel, E.M.: On the Externalization of Auditory Images. Presence Teleoper Virtual Env. 1, 251–257 (1992)
10. Majdak, P., Baumgartner, R., Jenny, C.: Formation of three-dimensional auditory space. In: The Technology of Binaural Understanding. Springer International Publishing (2020)
11. Kolarik, A.J., Moore, B.C.J., Zahorik, P., Cirstea, S., Pardhan, S.: Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. Atten. Percept. Psychophys. 78, 373–395 (2016). https://doi.org/10.3758/s13414-015-1015-1
12. Baumgartner, R., Majdak, P., Laback, B.: Modeling sound-source localization in sagittal planes for human listeners. J. Acoust. Soc. Am. 136, 791–802 (2014). https://doi.org/10.1121/1.4887447
13. Macpherson, E.A., Sabin, A.T.: Binaural weighting of monaural spectral cues for sound localization. J Acoust Soc Am. 121, 3677–3688 (2007). https://doi.org/10.1121/1.2722048
14. Van Opstal, A.J., Vliegen, J., Esch, T.V.: Reconstructing spectral cues for sound localization from responses to rippled noise stimuli. PLOS ONE. 12, e0174185 (2017). https://doi.org/10.1371/journal.pone.0174185
15. Strutt, J.W.L.R.: On our perception of sound direction. Philos Mag. 13, 214–232 (1907)
16. Macpherson, E.A., Middlebrooks, J.C.: Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. J. Acoust. Soc. Am. 111, 2219–2236 (2002). https://doi.org/10.1121/1.1471898
17. Hassager, H.G., Gran, F., Dau, T.: The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment. J. Acoust. Soc. Am. 139, 2992–3000 (2016). https://doi.org/10.1121/1.4950847
18. Shinn-Cunningham, B.G., Santarelli, S., Kopco, N.: Tori of confusion: binaural localization cues for sources within reach of a listener. J Acoust Soc Am. 107, 1627–36 (2000)
19. Devore, S., Ihlefeld, A., Hancock, K., Shinn-Cunningham, B., Delgutte, B.: Accurate Sound Localization in Reverberant Environments Is Mediated by Robust Encoding of

Spatial Cues in the Auditory Midbrain. Neuron. 62, 123–134 (2009). https://doi.org/10.1016/j.neuron.2009.02.018
20. Wood, K.C., Town, S.M., Bizley, J.K.: Neurons in primary auditory cortex represent sound source location in a cue-invariant manner. Nat. Commun. 10, 1–15 (2019). https://doi.org/10.1038/s41467-019-10868-9
21. Higgins, N.C., McLaughlin, S.A., Rinne, T., Stecker, G.C.: Evidence for cue-independent spatial representation in the human auditory cortex during active listening. Proc. Natl. Acad. Sci. 114, E7602–E7611 (2017). https://doi.org/10.1073/pnas.1707522114
22. Salminen, N.H., Takanen, M., Santala, O., Lamminsalo, J., Altoè, A., Pulkki, V.: Integrated processing of spatial cues in human auditory cortex. Hear. Res. 327, 143–152 (2015). https://doi.org/10.1016/j.heares.2015.06.006
23. Altmann, C.F., Terada, S., Kashino, M., Goto, K., Mima, T., Fukuyama, H., Furukawa, S.: Independent or integrated processing of interaural time and level differences in human auditory cortex? Hear. Res. 312, 121–127 (2014). https://doi.org/10.1016/j.heares.2014.03.009
24. Schröger, E.: Interaural time and level differences: Integrated or separated processing? Hear. Res. 96, 191–198 (1996). https://doi.org/10.1016/0378-5955(96)00066-4
25. Tardif, E., Murray, M.M., Meylan, R., Spierer, L., Clarke, S.: The spatio-temporal brain dynamics of processing and integrating sound localization cues in humans. Brain Res. 1092, 161–176 (2006). https://doi.org/10.1016/j.brainres.2006.03.095
26. Edmonds, B.A., Krumbholz, K.: Are Interaural Time and Level Differences Represented by Independent or Integrated Codes in the Human Auditory Cortex? J. Assoc. Res. Otolaryngol. 15, 103–114 (2014). https://doi.org/10.1007/s10162-013-0421-0
27. Colburn, H.S., Isabelle, S.K.: Models of Binaural Processing Based on Neural Patterns in the Medial Superior Olive. In: Cazals, Y., Horner, K., and Demany, L. (eds.) Auditory Physiology and Perception. pp. 539–545. Pergamon (1992)
28. Baldassi, S., Burr, D.C.: Feature-based integration of orientation signals in visual search. Vision Res. 40, 1293–1300 (2000). https://doi.org/10.1016/S0042-6989(00)00029-8
29. Thornton, M.A., Weaverdyck, M.E., Tamir, D.I.: The brain represents people as the mental states they habitually experience. Nat. Commun. 10, 1–10 (2019). https://doi.org/10.1038/s41467-019-10309-7
30. Palmer, J., Verghese, P., Pavel, M.: The psychophysics of visual search. Vision Res. 40, 1227–1268 (2000). https://doi.org/10.1016/S0042-6989(99)00244-8
31. Baldassi, S., Verghese, P.: Comparing integration rules in visual search. J. Vis. 2, 3–3 (2002). https://doi.org/10.1167/2.8.3
32. Song, Y.-H., Kim, J.-H., Jeong, H.-W., Choi, I., Jeong, D., Kim, K., Lee, S.-H.: A Neural Circuit for Auditory Dominance over Visual Perception. Neuron. 93, 940-954.e6 (2017). https://doi.org/10.1016/j.neuron.2017.01.006
33. Hartmann, W.M., Wittenberg, A.: On the externalization of sound images. J Acoust Soc Am. 99, 3678–88 (1996)
34. Baumgartner, R., Reed, D.K., Tóth, B., Best, V., Majdak, P., Colburn, H.S., Shinn-Cunningham, B.: Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias. Proc. Natl. Acad. Sci. 114, 9743–9748 (2017). https://doi.org/10.1073/pnas.1703247114
35. Boyd, A.W., Whitmer, W.M., Soraghan, J.J., Akeroyd, M.A.: Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers. J. Acoust. Soc. Am. 131, EL268–EL274 (2012). https://doi.org/10.1121/1.3687015
36. Colburn, H.S., Kulkarni, A.: Models of Sound Localization. In: Popper, A.N. and Fay, R.R. (eds.) Sound Source Localization. pp. 272–316. Springer New York (2005)
37. Zakarauskas, P., Cynader, M.S.: A computational theory of spectral cue localization. J Acoust Soc Am. 94, 1323–1331 (1993)
38. Baumgartner, R., Majdak, P., Laback, B.: Assessment of Sagittal-Plane Sound Localization Performance in Spatial-Audio Applications. In: The Technology of Binaural Listening. pp. 93–119. Springer, Berlin, Heidelberg (2013)

39. Baumgartner, R., Majdak, P., Laback, B.: Modeling the Effects of Sensorineural Hearing Loss on Sound Localization in the Median Plane. Trends Hear. 20, 1–11 (2016). https://doi.org/10.1177/2331216516662003
40. Romigh, G.D., Simpson, B.D., Iyer, N.: In Ear to Out There: A Magnitude Based Parameterization Scheme for Sound Source Externalization. Presented at the The 22nd International Conference on Auditory Display (ICAD–2016) , Canberra, Australia July 2 (2016)
41. Georganti, E., May, T., Par, S. van de, Mourjopoulos, J.: Sound Source Distance Estimation in Rooms based on Statistical Properties of Binaural Signals. IEEE Trans. Audio Speech Lang. Process. 21, 1727–1741 (2013). https://doi.org/10.1109/TASL.2013.2260155
42. Leclère, T., Lavandier, M., Perrin, F.: On the externalization of sound sources with headphones without reference to a real source. J. Acoust. Soc. Am. 146, 2309–2320 (2019). https://doi.org/10.1121/1.5128325
43. Mills, A.W.: On the Minimum Audible Angle. J. Acoust. Soc. Am. 30, 237–246 (1958). https://doi.org/10.1121/1.1909553
44. Lyon, R.F.: All pole models of auditory filtering. In: E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and Hecht-Poinar, E. (eds.) Diversity in auditory mechanics. pp. 205–211. World Scientific Publishing, Singapore (1997)
45. Baumgartner, R., Majdak, P., Laback, B.: Erratum: Modeling sound-source localization in sagittal planes for human listeners [J. Acoust. Soc. Am. 136, 791–802 (2014)]. J. Acoust. Soc. Am. 140, 2456–2456 (2016). https://doi.org/10.1121/1.4964753
46. Zilany, M.S.A., Bruce, I.C., Carney, L.H.: Updated parameters and expanded simulation options for a model of the auditory periphery. J. Acoust. Soc. Am. 135, 283–286 (2014). https://doi.org/10.1121/1.4837815
47. Katz, B.F.G., Noisternig, M.: A comparative study of interaural time delay estimation methods. J. Acoust. Soc. Am. 135, 3530–3540 (2014). https://doi.org/10.1121/1.4875714
48. Reiss, L.A.J., Young, E.D.: Spectral edge sensitivity in neural circuits of the dorsal cochlear nucleus. J. Neurosci. 25, 3680–3691 (2005). https://doi.org/10.1523/JNEUROSCI.4963-04.2005
49. May, B.J.: Role of the dorsal cochlear nucleus in the sound localization behavior of cats. Hear. Res. 148, 74–87 (2000). https://doi.org/10.1016/S0378-5955(00)00142-8
50. Spagnol, S.: On distance dependence of pinna spectral patterns in head-related transfer functions. J. Acoust. Soc. Am. 137, EL58–EL64 (2015). https://doi.org/10.1121/1.4903919
51. Brungart, D.S., Rabinowitz, W.M.: Auditory localization of nearby sources. Head-related transfer functions. J. Acoust. Soc. Am. 106, 1465–1479 (1999). https://doi.org/10.1121/1.427180
52. Macpherson, E.A., Middlebrooks, J.C.: Vertical-plane sound localization probed with ripple-spectrum noise. J Acoust Soc Am. 114, 430–445 (2003). https://doi.org/10.1121/1.1582174
53. Denham, S.L., Winkler, I.: Predictive coding in auditory perception: challenges and unresolved questions. Eur. J. Neurosci. 51, 1151–1160 (2020). https://doi.org/10.1111/ejn.13802
54. Klein, F., Werner, S., Mayenfels, T.: Influences of Training on Externalization of Binaural Synthesis in Situations of Room Divergence. J. Audio Eng. Soc. 65, 178–187 (2017)
55. Li, S., Baumgartner, R., Peissig, J.: Modeling perceived externalization of a static, lateral sound image. Acta Acust. 4, 21 (2020). https://doi.org/10.1051/aacus/2020020
56. Brimijoin, W.O., Boyd, A.W., Akeroyd, M.A.: The Contribution of Head Movement to the Externalization and Internalization of Sounds. PLoS ONE. 8, e83068 (2013). https://doi.org/10.1371/journal.pone.0083068
57. Li, S., Schlieper, R., Peissig, J.: The Role of Reverberation and Magnitude Spectra of Direct Parts in Contralateral and Ipsilateral Ear Signals on Perceived Externalization. Appl. Sci. 9, 460 (2019). https://doi.org/10.3390/app9030460

58. Hendrickx, E., Stitt, P., Messonnier, J.-C., Lyzwa, J.-M., Katz, B.F.G., Boishéraud, C. de: Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis. J. Acoust. Soc. Am. 141, 2011–2023 (2017). https://doi.org/10.1121/1.4978612
59. Kates, J.M., Arehart, K.H., Muralimanohar, R.K., Sommerfeldt, K.: Externalization of remote microphone signals using a structural binaural model of the head and pinna. J. Acoust. Soc. Am. 143, 2666–2677 (2018). https://doi.org/10.1121/1.5032326
60. Mendonça, C., Campos, G., Dias, P., Santos, J.A.: Learning Auditory Space: Generalization and Long-Term Effects. PLOS ONE. 8, e77900 (2013). https://doi.org/10.1371/journal.pone.0077900
61. Hendrickx, E., Stitt, P., Messonnier, J.-C., Lyzwa, J.-M., Katz, B., de Boishéraud, C.: Improvement of Externalization by Listener and Source Movement Using a “Binauralized” Microphone Array. J. Audio Eng. Soc. 65, 589–599 (2017). https://doi.org/10.17743/jaes.2017.0018
62. Reijniers, J., Vanderelst, D., Jin, C., Carlile, S., Peremans, H.: An Ideal-observer Model of Human Sound Localization. Biol Cybern. 108, 169–181 (2014). https://doi.org/10.1007/s00422-014-0588-4
63. Marelli, D., Baumgartner, R., Majdak, P.: Efficient Approximation of Head-Related Transfer Functions in Subbands for Accurate Sound Localization. IEEEACM Trans. Audio Speech Lang. Process. 23, 1130–1143 (2015). https://doi.org/10.1109/TASLP.2015.2425219
64. Simon, L.S.R., Zacharov, N., Katz, B.F.G.: Perceptual attributes for the comparison of head-related transfer functions. J. Acoust. Soc. Am. 140, 3623–3632 (2016). https://doi.org/10.1121/1.4966115
65. Crawford, S., Audfray, R., Jot, J.-M.: Quantifying HRTF Spectral Magnitude Precision in Spatial Computing Applications. Presented at the Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality August 13 (2020)