Decision making in auditory externalization perception

Robert Baumgartner[1] and Piotr Majdak

Acoustics Research Institute, Austrian Academy of Sciences

Abstract

Under natural listening conditions, humans perceptually attribute a sound to an external object located in their environment. This so-called auditory externalization is often distorted when the sound is produced via hearing devices such as headphones or hearing aids, resulting in sources being perceived inside the head. Psychoacoustic studies suggest a mixed role of various cues contributing to the externalization process. As such, we developed a model framework able to probe the contribution of those cues and test the perceptual decision strategy of listeners performing externalization tasks. Inspired by investigations in vision, we contrasted different decision strategies including dynamic and static approaches. Our results show that the decision strategy following a static combination of cues best explains perceived externalization in various listening situations. Our analysis indicates that listeners mainly base their perceptual decision on both monaural and interaural spectral shapes, without a significant contribution of other auditory cues. This provides support for predictive processing in spatial hearing and its precision-weighted error accumulation in perceptual inference. From a technical perspective, the proposed model predicts auditory externalization under the most fragile condition of static listening and has the potential to significantly advance the future development of spatially-aware hearing technologies by helping to identify perceptually relevant auditory distortions.

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 (Friston, 2005; Hawkins & Blakeslee, 2004; Walsh et al., 2020). These theories convince by examples coming from various areas of sensory perception and higher-order cognitive functions (Den Ouden et al., 2012). In the context of auditory perception, temporal estimates are often based on short-term regularities (Baldeweg, 2006; Ulanovsky et al., 2003) while long-term expectations are often crucial for spatial tasks (Majdak et al., 2020; van der Heijden et al., 2019). 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 (Gardner, 2019) but remained an unresolved topic in spatial hearing (van der Heijden et al., 2019). 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; Loomis, 1992), which denotes the lack to associate sensations with external objects (Weber, 1848).

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 (Blauert, 1997; Durlach et al., 1992; Majdak et al., 2020). 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. Many studies aimed to identify in particular spectral cues affecting the perceptual externalization of sound sources (Baumgartner et al., 2017; Boyd et al., 2012; Hartmann & Wittenberg, 1996; Hassager et al., 2016; Zhang & Hartmann, 2010). These cues can be grouped into two categories (Tab. 1): one category poses monaural cues, in line with studies on sound localisation beyond the horizontal plane suggesting that the auditory system processes spectral cues within monaural pathways (Baumgartner et al., 2014; Macpherson & Sabin, 2007; May, 2000; Slattery III & Middlebrooks, 1994), and the

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.

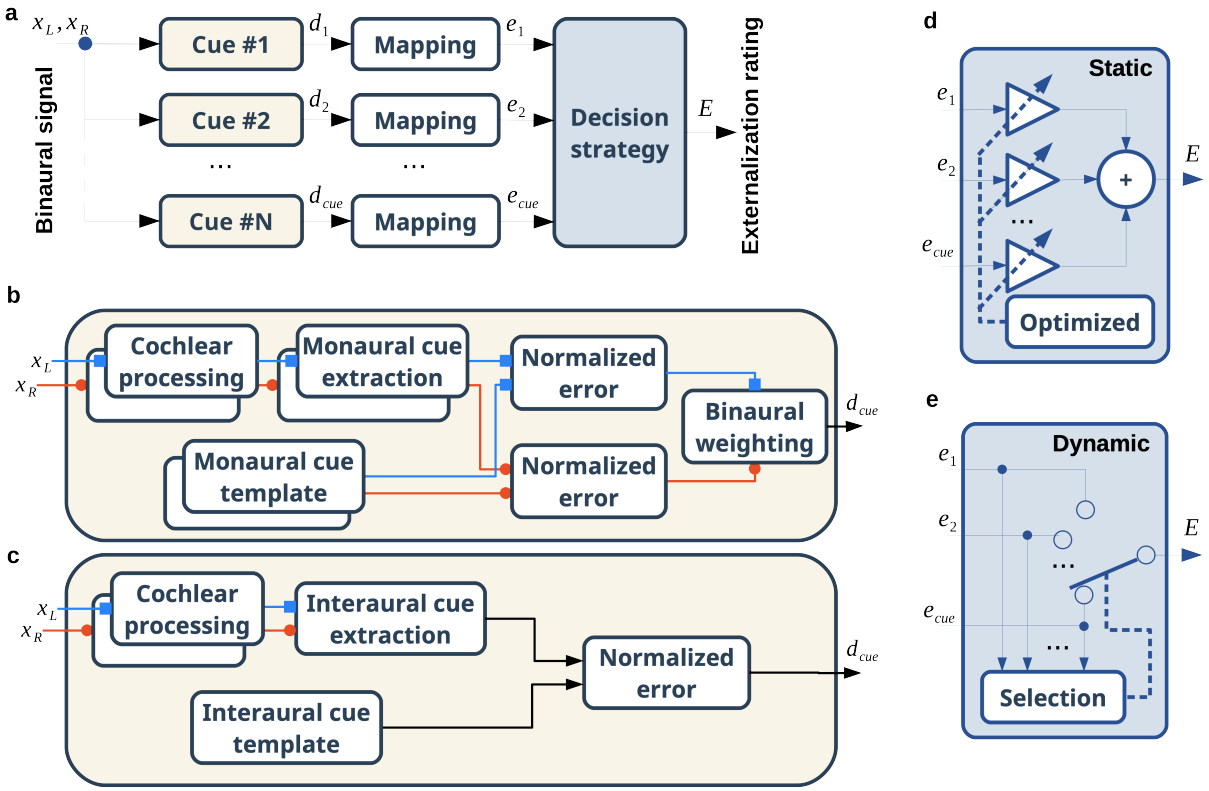

Fig. 1 | General structure of the sound externalization framework. ***a****, Processing of the binaural signal* x_L, x_R *resulting in a cue-based error* d_{cue}*, a stage mapping the cue-based error to a normalized externalization rating* e_{cue}*, and a decision stage combining the cue-specific ratings 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).*

other category poses interaural cues, such as an interaural time difference (ITD) or interaural intensity difference (IID), which are well known to be crucial for sound localisation within the horizontal plane (Macpherson & Middlebrooks, 2002; Strutt, 1907) but may also affect spatial hearing in other dimensions (Hassager et al., 2016; Shinn-Cunningham et al., 2000). Besides the multitude of potential cues, a general problem is 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 (Devore et al., 2009), 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 (Higgins et al., 2017; Salminen et al., 2015; Wood et al., 2019), but also retain independent information about spatial cues (Altmann et al., 2014; Edmonds & Krumbholz, 2014; Schröger, 1996; Tardif et al., 2006), allowing the system to decide how likely they would have arisen from one or multiple sources (Edmonds & Krumbholz, 2014), 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 weighing 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 (Colburn & Isabelle, 1992) over higher cortical levels integrating visual orientation cues (Baldassi & Burr, 2000) to even more abstract levels such as the representation of person identities (Thornton et al., 2019). 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 in the context of visual search (Baldassi & Verghese, 2002; Palmer et al., 2000) or audio-visual dominance (Song et al., 2017). 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. The 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. We used that framework to simulate five experiments from representative psychoacoustic externalization studies

(Baumgartner et al., 2017; Boyd et al., 2012; Hartmann & Wittenberg, 1996; Hassager et al., 2016), all focusing on the effects of spectral distortions under static listening conditions, but differing considerably in manipulation method and bandwidth. The listener's expectations for perceptual inference are represented as internal templates and used for comparison with incoming cue-specific information, distinguishing between processing monaural and interaural cues (Fig. 1b,c). The decision strategies were not only compared for all those cues, but also for a successively reduced set of cues, in order to address the potential redundancy between them. Our results reveal that auditory externalization judgments under situations with constant reverberation underlie a fixed weighting scheme between monaural and interaural spectral cues, rather than a dynamic selection of those auditory cues.

Results

Individual cues: large differences across experimental conditions

We first investigated the ability of each individual cue to predict the effects of signal manipulations on auditory externalization as tested in five previous headphone experiments (Baumgartner et al., 2017; Boyd et al., 2012; Hartmann & Wittenberg, 1996; Hassager et al., 2016). Figure 2 shows the predicted externalization ratings together with the actual ratings replotted from the corresponding experiments with normalized rating scales. The model's normalization stage (Fig. 1b,c) requires a listener-specific sensitivity (Baumgartner et al., 2016; Majdak et al., 2014), which was fitted to optimally predict the externalization ratings separately for each experiment and cue (Fig. 2e).

Exp. I: Effect of IID modifications at low frequencies

The first experiment presented listeners with harmonic tone complexes from a mid-left position (Fig. 2a, Exp. I; Hartmann & Wittenberg, 1996). The magnitudes of all harmonics up to a certain harmonic were set to the interaural average, effectively removing IIDs up to that harmonic's frequency (see Methods for further details).

The monaural spectral cues (MSS and MSSD) showed a non-monotonic relationship between the modified frequency range and predicted ratings, which largely diverged from the actual ratings especially for the largest frequency ranges and resulted in large prediction 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 predict 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 predict 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). Predictions 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 prediction error of 0.06.

Exp. II: Effect of monaural modifications at low frequencies

The second experiment tested the same stimulus as in Exp. II, but instead of removing interaural differences, the ipsilateral magnitude spectrum was flattened up to a certain har-

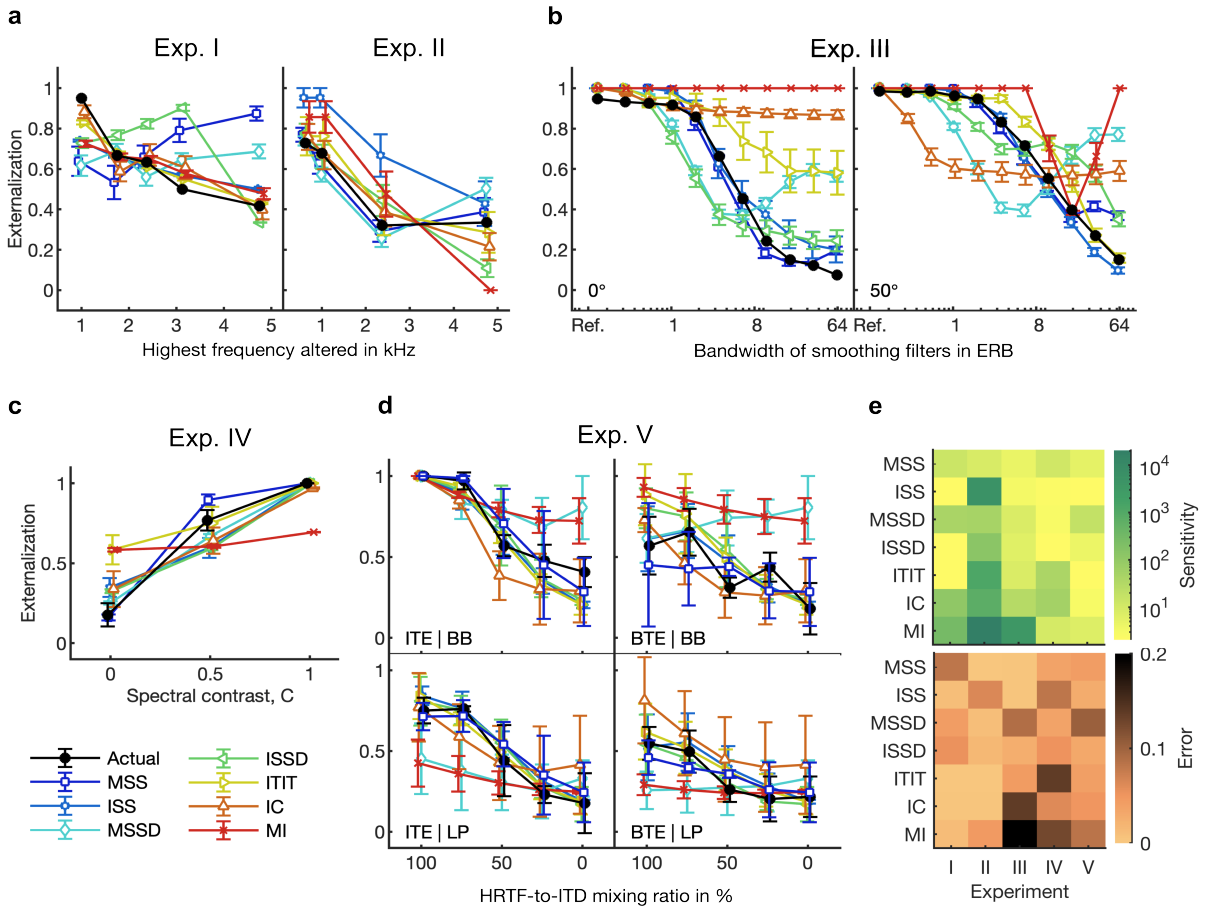

Fig. 2 | Externalization ratings: actual data from psychoacoustic experiments *(closed circles)* ***and predictions from the single-cue models*** *(open symbols).* ***a****, Effects of low-frequency modifications tested by Hartmann and Wittenberg (1996). Exp. I: IID set to zero (bilateral average magnitude); actual data from their Fig. 7. Exp. II: ipsilateral magnitudes flattened (IID compensated by contralateral magnitudes); actual data from their Fig. 8.* ***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. (2016).* ***c****, Exp. IV: effect of spectral smoothing in high frequencies as a function of spectral contrast (C=1: natural listener-specific spectral coloration; C=0: flat spectra); actual data calculated from the paired-comparison data from Baumgartner et al. (2017).* ***d****, Exp. V: effects of stimulus bandwidth and microphone casing for various mixes between simulations based on listener-specific binaural room impulse responses (100%) and time-delay stereophony (0%); actual data from Boyd et al. (2012). ITE: in-the-ear casing; BTE: behind-the-ear casing; BB: broadband stimulus; LP: low-pass filtered at 6.5 kHz; HRTF: head-related transfer function (spectral coloration due to directional acoustic filtering).* ***e****, Cue-specific sensitivities and RMS prediction errors revealed as best explaining the actual externalization ratings. Per experiment, the smallest error indicates the most informative cue.*

monic while the contralateral magnitudes were shifted in parallel (Hartmann & Wittenberg, 1996). 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).

In contrast to Exp. I, the predictions based on monaural spectral cues (MSS and MSSD) reflected the decrease in actual ratings very well, as indicated by RMS prediction error as low as 0.05 (for MSS). As expected, predictions 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 third experiment investigated the effect of spectral smoothing on the auditory externalization of noise stimuli band-limited up to 6 kHz and presented sounds either from the front (0° azimuth) or from the left (50° azimuth; Hassager et al., 2016). Spectral smoothing was implemented by applying filters of various equivalent rectangular bandwidths (ERBs), with broader filters removing more spectral details. This procedure caused a degradation in perceived externalization, which was more gradual for the lateral as compared to the frontal direction (Fig. 2b).

Again, only some of the cues were able to predict 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 much better predictions (errors of 0.18 and 0.25, respectively). Both cues yielded predictions 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; Hassager et al., 2016). 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

The fourth experiment also tested the effect of spectral smoothing (Baumgartner et al., 2017), however, focusing on the high-frequency range from 1 kHz to 16 kHz, where the pinnae induce the most significant directional spectral variations. There, listeners were presented to noise stimuli being spectrally flat (C = 0), spectrally shaped according to their individual acoustics (C = 1), or something in between (C = 0.5). Flattening the spectral contrast concordantly degraded externalization perception (Fig. 2c).

Broadband cues (ITIT and MI) were hardly affected by this method and were not able to explain the actual results. Best predictions 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 predictions 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

The last experiment presented broadband speech samples from a mid-left position and compared externalization ratings for different sensor positions, with sensors placed in the ear (ITE) distorting natural spectral cues less than behind the ear (BTE), stimulus bandwidths, and mixing ratios with stereophonic recordings providing only an ITD but no IID. The predictions showed much variability (Fig. 2d) across cues and conditions. The broadband BTE condition caused the most distinct spectral modifications and was particularly informative about the predictive power of considered cues. Most cues were able to follow the actual ratings quite well (except MSSD and MI). On average across conditions, expectations based on interaural spectral templates (ISS, prediction error of 0.16) seem to have been most relevant to the listeners in this experiment.

Individual cue summary

The overall picture of prediction errors suggests that there is no single cue able to predict externalization for all experiments but that there is a small set of cues which listeners 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.

The monaural cue (MSS) yielded best predictions in three out of five experiments as well as on average across the five experiments (prediction 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 (Reiss & Young, 2005). 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 prediction 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 many 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 errors.

In summary, we found considerable variance in cue-specific predictions. Predictions 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 predictions based on that particular would fail in situations similar to Exp. I. This indicates that the auditory system does not simply rely on a single cue and rises the 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 predictions (Fig. 1b). The weights were obtained from an optimization procedure minimizing the prediction errors. The dynamic decision strategies, that is, WTA, MTA, and LTA, selected the largest, median, and smallest predicted externalization rating, respectively, obtained across the considered cues. By doing so, the cue weights

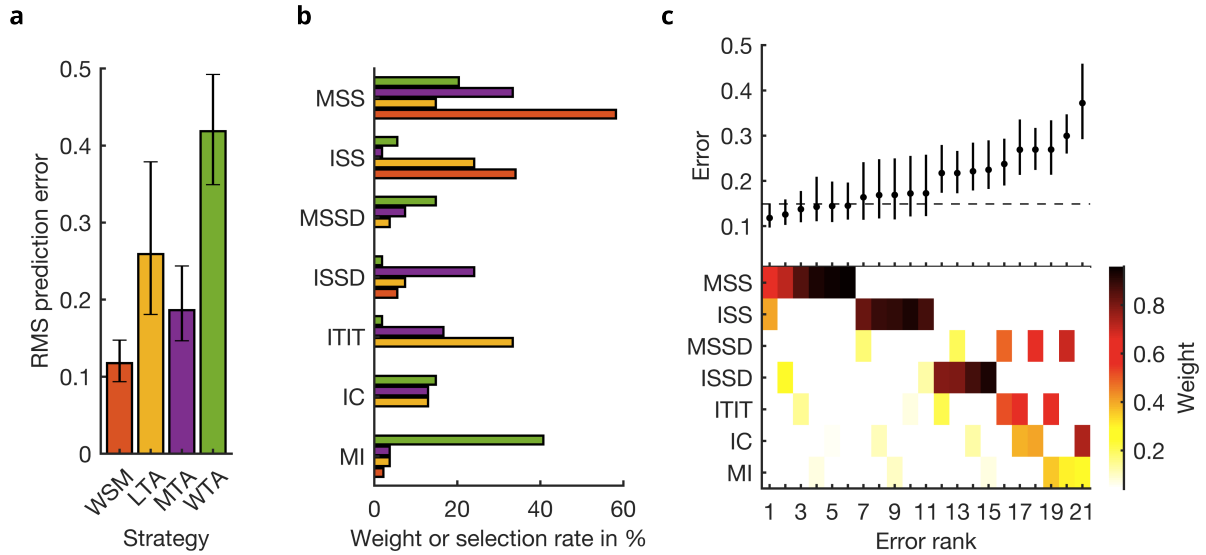

Fig. 3 | Prediction errors for different decision strategies and cue combinations show that static combination (WSM) based on monaural and interaural spectral shape cues (MSS, ISS) performs best. a, *Prediction errors for different strategies and pooled experimental data. Error bars denote 95% confidence intervals estimated via boot strapping. WSM: weighted-sum model; L/M/WTA: looser/median/winner takes all.* ***b,*** *Individual cue contributions. Cue abbreviations defined in Tab. 1.* ***c, Top:*** *Prediction 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.*

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 prediction errors (Fig. 3a). It predicted 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 marginally only (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 (Hassager et al., 2016). 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 perfectionist approach (WTA) performed particularly poorly, being close to the chance rate of 0.5. The cues selected by the 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 prediction results where this particular cue also provided good results for some conditions. Concordantly, the large variance in prediction 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 prediction error, are shown in Fig. 3c. As expected there is a considerable variance in the prediction 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 prediction 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 prediction error remained small regardless of the other considered cue. As long the that cue was involved, the prediction errors did not increase significantly (remained within the confidence intervals of the best condition). Thus, the monaural spectral shape cue seems to be the most important cue for auditory ex-

ternalization 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 prediction errors significantly increased ($p < 0.05$, compared to the best condition) as soon as neither the monaural (MSS) nor the interaural spectral shape cue (ISS) 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 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. Contrary to results derived from previous investigations in vision (Baldassi & Verghese, 2002; Palmer et al., 2000; Song et al., 2017), our results suggest that a static, weighted combination (WSM) rather than a dynamic selection (LTA, MTA, WTA) of multiple cues drives perceptual decisions on 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.

Predictive processing

To explain these results, we may need to consider the highly predictable world we live in. According to theories of predictive processing, such an environment allows us to build internal models preparing the sensory systems for processing and by doing so, we can react faster and more accurately to incoming stimuli. Predictive processing assumes 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 (Walsh et al., 2020). 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. In fact, the perceptual weights we obtained are well in line with this consideration because in the acoustic near-field of a listener both monaural spectral cues (MSS; Spagnol, 2015) and interaural spectral cues (ISS; Brungart & Rabinowitz, 1999) are known to be informative about absolute distance. The little importance of other cues in our model is also in line with this reasoning as monaural intensities, for instance, inform distance inference only on a relative scale with intrinsic ambiguity.

Furthermore, the perceptual inference underlying auditory externalization is known to be adaptive (Gil-Carvajal et al., 2016; Mendonça et al., 2013). This adaptation could be grounded on changes in expectations in line with predictive processing implemented in a straight forward fashion, incorporate dynamic strategies to resolve ambiguous sensory information, or consist of a mixture of both. Our findings do not support the existence of adaptive strategies of cue selection but promote a straight forward implementation of predictive processing. Hence, adaptivity in auditory externalization seems to be grounded solely on the generation of new expectations, i.e., updating of templates. Such template updates were, however, not considered in our model simulations because all simulated experiments were conducted under static listening conditions with a clear individualized and well externalized reference condition.

Dominant cues

A closer look at the two major cues favoured by our model selection procedure, that is, monaural and interaural spectral shapes, shows that those two cues are also well-known to be important for and complementary in the process of sound localisation. Monaural spectral shapes comprise localisation cues within sagittal planes, i.e., including both the vertical and front/back dimension (Baumgartner et al., 2013). The current understanding of this cue is that it resembles the processing of monaural positive spectral gradient profiles (MSS). In line with previous assumptions (Baumgartner et al., 2014, 2016) and electrophysiological measurements in the dorsal cochlear nucleus (Reiss & Young, 2005), limiting the contribution of negative spectral gradients here turned out to improve our predictions of perceived externalization.

The second cue comprises the spectral shape of interaural intensity differences (ISS) and is very well established as a cue for auditory lateralization (Macpherson & Middlebrooks, 2002; Strutt, 1907). Given the inherent frequency selectivity of the auditory periphery it is reasonable to believe that relative differences to an internal reference can be evaluated on a frequency-specific level (Hassager et al., 2016), as implemented in the ISS cue of our model. The accessibility or predictability of such interaural spectral evaluation must, however, be limited by the auditory system to some degree because otherwise it would be hard to argue why the monaural counterpart that suffers from ambiguity with the source spectrum (Macpherson & Middlebrooks, 2003) is considered for spatial inference at all.

Directions for future investigations

The strong overlap of cues for the perceptual externalization and directional localisation of sounds suggests also a strong overlap of underlying perceptual processes. However, the externalization model proposed here does not explicitly consider interactions with directional sound localisation (Brimijoin et al., 2013; Hendrickx, Stitt, Messonnier, Lyzwa, Katz, & Boishéraud, 2017; Kates et al., 2018; Li et al., 2019). In our simulations, template comparis-

ons were only performed for the reference direction, ignoring that there might be a strong match to a template from another direction yielding strong externalization. For example, spectral cue changes may have elicited changes in the perceived polar angle within a sagittal plane but such perceptual changes were not controlled. However, except for one study (Baumgartner et al., 2017), listeners in the experimental paradigms were asked to rate 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 would allow to consider this interdependency in model-based analysis. Future psychoacoustic experiments specifically targeting the spectral processing asymmetry proposed by our model investigations also appear important because conclusions drawn from existing data are limited by the fact that positive and negative spectral gradients were affected quite equally and simultaneously in most conditions.

Reverberation is usually omnipresent; it smears the spectral profile and decorrelates the binaural signal, effectively increasing the variance of interaural cues and affecting the reliability of cues. The degree of externalization is known to increase with the amount of reverberation as long as the related cues are consistent with the listener's expectations about the room acoustics (Klein et al., 2017). Hence, also the evaluation of reverberation-related cues appears consistent with our template-based modelling framework, which thus may also help resolve the similarly debated nature of those cues (Catic et al., 2015; Leclère et al., 2019; Li et al., 2018) in future analyses.

A more substantial extension of the promoted 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 (Brimijoin et al., 2013; Hendrickx, Stitt, Messonnier, Lyzwa, Katz, & Boishéraud, 2017; Hendrickx, Stitt, Messonnier, Lyzwa, Katz, & de Boishéraud, 2017). Internal generative models are considered to constantly shape the listener's expectations by means of either explicit or implicit predictive coding (Denham & Winkler, 2020). Embedding our proposed framework and included decision weights in a larger predictive coding framework 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 model successfully predicts auditory externalization ratings in static listening conditions without prior knowledge about the experiment – a feature often requested by scientists from the area of psychophysics and neuroscience, as well as engineers working on audio and human interfaces such as augmented and virtual reality. 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. To simplify the application and further development of the model, and guarantee repro-

ducibility of results, implementations of both the model and the model simulations are publicly available as part of the Auditory Modelling Toolbox (model named baumgartner2020; Majdak & Co., 2020; Søndergaard & Majdak, 2013).

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 used a binaural signal as an input and predicted externalization ratings as an output after performing comparisons with an 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.

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; Baumgartner et al., 2014) and interaural spectral shape (ISS; Hassager et al., 2016), we considered various cues that could be associated with externalization or distance perception of anechoic sounds: the difference in broadband monaural intensity (MI), the spectral standard deviation of IIDs (ISSD; Georganti et al., 2013), the broadband interaural coherence (IC; Leclère et al., 2019), and the inconsistency between ITD and IID (ITIT). In analogy to the ISSD as a direct measure of ISS, we also introduced a direct measure assessing the spectral standard deviation of spectral gradients (MSSD).

For each cue, we first calculated an internal error metric relating an acoustic signal feature with externalization. The error metric was then mapped to an externalization rating, which has been scaled individually for each experiment. Thus, both the mapping and calculation of error metrics depended on the considered cue and 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 $\Delta MI = \left| MI_{target} - MI_{template} \right|$ 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 (Mills, 1958) 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 $IC = \lim_{\tau} \inf \int x_L(t-\tau)\, x_R(t)\, dt$ for the binaural signals x_L and $\pm\pi/2$ 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:

$$d_{IC}=\frac{|IC_{target}-IC_{template}|}{IC_{template}} \quad (2)$$

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 (Lyon, 1997). The spectral excitation profiles were computed from the logarithm of the RMS energy within every frequency band (Baumgartner et al., 2014). Audibility thresholds for band-limited signals were roughly approximated by generally considering stimulus SPLs to be 70 dB and a within-band threshold of 20 dB. Assuming stationary input signals, the spectral profiles were averaged over time, yielding spectral profiles as a function of frequency band, $p(f)$. Further, the interaural difference of the spectral profiles yielded IIDs as a function of frequency band, $ILD(f)$.

For the ISSD cue, the model evaluated standard deviations (SD) of $ILD(f)$ and computed the negative difference of these deviations between the target and template, relative to the template deviation, yielding:

$$d_{ISSD}=\frac{|SD_f(ILD_{target}(f))-SD_f(ILD_{template}(f))|}{SD_f(ILD_{template}(f))} \quad (3)$$

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:

$$d_{ISS}=\frac{1}{N_f}\sum_f \frac{\Delta|ILD(f)|}{|ILD_{template}(f)|}, \quad (4)$$

with N_f being the number of frequency bands.

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

$$d_{ITIT}=\left|\frac{\Delta ITD}{ITD_{template}}-\frac{\Delta ILD}{ILD_{template}}\right|, \quad (5)$$

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 (Katz & Noisternig, 2014).

For the MSS and MSSD cues, positive spectral gradient profiles were derived exactly as in Baumgartner et al. (2016). 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, yielding

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

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 (Baumgartner et al., 2014; Macpherson & Sabin, 2007), 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 deviation metrics 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. Mapping sensitivity is denoted by $1/S_{cue}$ because a larger $1/S_{cue}$ yields a steeper mapping function that projects smaller errors to smaller externalization ratings. This cue- and experiment-specific sensitivity was obtained by minimizing the squared prediction error, which was defined as the RMS of differences between actual and predicted 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, the optimal weights were obtained by optimizing weights scaling the contribution of a cue to the ratings, i.e., minimizing the prediction 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 predictions, and the individual summation weights were optimized to best fit the data of all experiments.

Considered studies

For the modelling, we considered results of five experiments (Baumgartner et al., 2017; Boyd et al., 2012; Hartmann & Wittenberg, 1996; Hassager et al., 2016). The pool of experiments was a compromise of data availability and the degree to which the tests isolate specific cues. The experiments differed in whether they provided visual information about the reference source and tested in anechoic or reverberant environments. While the constant reverberation and visual information may or may not have stabilized auditory externalization, they certainly did not prevent the tested signal modifications to be effective within the tested condition. In our study, we thus assumed that such differences in experimental procedures do not modulate our effects of interest.

Exp. I and II: Hartmann and Wittenberg (1996) 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 was presented via headphones and filtered with individualized HRTFs corresponding to 37° azimuth (0° elevation). For modelling, we used HRTFs from 21 exemplary listeners contained in the ARI database.

Exp. III: Hassager et al. (2016) investigated the effect of spectral smoothing on auditory externalization of Gaussian white low-frequency noises, i.e., band-limited from 50 Hz to 6000 Hz. 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 smooth 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. Because the original BRIRs were not accessible, our model predictions were again based on the same 21

(anechoic) HRTFs from the ARI database and only addressed the effect of spectrally smoothing the direct path.

Exp. IV: Absolute externalization ratings from Baumgartner et al. (2017) were estimated from the listeners' paired binary judgments by calculating mean relative frequencies of "farther" judgments. We evaluated only data from their "discontinuous trial" condition of their Exp. II, which did not elicit an auditory looming bias, and from the ten out of twelve listeners who perceived the spectrally flat sound (C = 0) as being closer than the individualized reference sound (C = 1).

Exp. V: Boyd et al. (2012) used individualized BRIRs to simulate a talker positioned at 30° azimuth. They compared externalization ratings for in-the-ear (ITE) and behind-the-ear (BTE) microphone casings as well as broadband (BB) and 6.5-kHz-low-pass (LP) filtered stimuli at various mixing ratios with stereophonic recordings providing only an ITD but no IID. 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.

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.) and Facebook Reality Labs.

Author Contributions

R.B. designed the research and performed the simulations; R.B. and P.M. interpreted the results and wrote the manuscript.

Competing Interests

The authors declare no competing interests.

References

Altmann, C. F., Terada, S., Kashino, M., Goto, K., Mima, T., Fukuyama, H., & Furukawa, S. (2014). Independent or integrated processing of interaural time and level differences in human auditory cortex? *Hearing Research*, *312*, 121–127. https://doi.org/10.1016/j.heares.2014.03.009

Baldassi, S., & Burr, D. C. (2000). Feature-based integration of orientation signals in visual search. *Vision Research*, *40*(10), 1293–1300. https://doi.org/10.1016/S0042-6989(00)00029-8

Baldassi, S., & Verghese, P. (2002). Comparing integration rules in visual search. *Journal of Vision*, *2*(8), 3–3. https://doi.org/10.1167/2.8.3

Baldeweg, T. (2006). Repetition effects to sounds: Evidence for predictive coding in the auditory system. *Trends in Cognitive Sciences*, *10*(3), 93–94. https://doi.org/10.1016/j.tics.2006.01.010

Baumgartner, R., Majdak, P., & Laback, B. (2013). Assessment of Sagittal-Plane Sound Localization Performance in Spatial-Audio Applications. In *The Technology of Binaural Listening* (pp. 93–119). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37762-4_4

Baumgartner, R., Majdak, P., & Laback, B. (2014). Modeling sound-source localization in sagittal planes for human listeners. *The Journal of the Acoustical Society of America*, *136*(2), 791–802. https://doi.org/10.1121/1.4887447

Baumgartner, R., Majdak, P., & Laback, B. (2016). Modeling the Effects of Sensorineural Hearing Loss on Sound Localization in the Median Plane. *Trends in Hearing*, *20*, 1–11. https://doi.org/10.1177/2331216516662003

Baumgartner, R., Reed, D. K., Tóth, B., Best, V., Majdak, P., Colburn, H. S., & Shinn-Cunningham, B. (2017). Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias. *Proceedings of the National Academy of Sciences*, *114*(36), 9743–9748. https://doi.org/10.1073/pnas.1703247114

Blauert, J. (1997). *Spatial hearing. The Psychophysics of Human Sound Localization* (2nd edition). MIT-Press.

Boyd, A. W., Whitmer, W. M., Soraghan, J. J., & Akeroyd, M. A. (2012). Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers. *The Journal of the Acoustical Society of America*, *131*(3), EL268–EL274. https://doi.org/10.1121/1.3687015

Brimijoin, W. O., Boyd, A. W., & Akeroyd, M. A. (2013). The Contribution of Head Movement to the Externalization and Internalization of Sounds. *PLoS ONE*, *8*(12), e83068. https://doi.org/10.1371/journal.pone.0083068

Brungart, D. S., & Rabinowitz, W. M. (1999). Auditory localization of nearby sources. Head-related transfer functions. *The Journal of the Acoustical Society of America*, *106*(3), 1465–1479. https://doi.org/10.1121/1.427180

Catic, J., Santurette, S., & Dau, T. (2015). The role of reverberation-related binaural cues in the externalization of speech. *The Journal of the Acoustical Society of America*, *138*(2), 1154–1167. https://doi.org/10.1121/1.4928132

Colburn, H. S., & Isabelle, S. K. (1992). Models of Binaural Processing Based on Neural Patterns in the Medial Superior Olive. In Y. Cazals, K. Horner, & L. Demany (Eds.), *Auditory Physiology and Perception* (pp. 539–545). Pergamon. https://doi.org/10.1016/B978-0-08-041847-6.50066-6

Den Ouden, H. E., Kok, P., & De Lange, F. P. (2012). How Prediction Errors Shape Perception, Attention, and Motivation. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00548

Denham, S. L., & Winkler, I. (2020). Predictive coding in auditory perception: Challenges and unresolved questions. *European Journal of Neuroscience*, *51*(5), 1151–1160. https://doi.org/10.1111/ejn.13802

Devore, S., Ihlefeld, A., Hancock, K., Shinn-Cunningham, B., & Delgutte, B. (2009). Accurate Sound Localization in Reverberant Environments Is Mediated by Robust Encoding of Spatial Cues in the Auditory Midbrain. *Neuron*, *62*(1), 123–134. https://doi.org/10.1016/j.neuron.2009.02.018

Durlach, N. I., Rigopulos, A., Pang, X. D., Woods, W. S., Kulkarni, A., Colburn, H. S., & Wenzel, E. M. (1992). On the Externalization of Auditory Images. *Presence: Teleoper. Virtual Environ.*, *1*(2), 251–257.

Edmonds, B. A., & Krumbholz, K. (2014). Are Interaural Time and Level Differences Represented by Independent or Integrated Codes in the Human Auditory Cortex? *Journal of the Association for Research in Otolaryngology*, *15*(1), 103–114. https://doi.org/10.1007/s10162-013-0421-0

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Gardner, J. L. (2019). Optimality and heuristics in perceptual neuroscience. *Nature Neuroscience*, *22*(4), 514–523. https://doi.org/10.1038/s41593-019-0340-4

Georganti, E., May, T., Par, S. van de, & Mourjopoulos, J. (2013). Sound Source Distance Estimation in Rooms based on Statistical Properties of Binaural Signals. *IEEE Transactions on Audio, Speech, and Language Processing*, *21*(8), 1727–1741. https://doi.org/10.1109/TASL.2013.2260155

Gil-Carvajal, J. C., Cubick, J., Santurette, S., & Dau, T. (2016). Spatial Hearing with Incongruent Visual or Auditory Room Cues. *Scientific Reports*, *6*, 37342. https://doi.org/10.1038/srep37342

Hartmann, W. M., & Wittenberg, A. (1996). On the externalization of sound images. *J Acoust Soc Am*, *99*(6), 3678–3688.

Hassager, H. G., Gran, F., & Dau, T. (2016). The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment. *The Journal of the Acoustical Society of America*, *139*(5), 2992–3000. https://doi.org/10.1121/1.4950847

Hawkins, J., & Blakeslee, S. (2004). *On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines*. Times Books.

Hendrickx, E., Stitt, P., Messonnier, J.-C., Lyzwa, J.-M., Katz, B., & de Boishéraud, C. (2017). Improvement of Externalization by Listener and Source Movement Using a

“Binauralized” Microphone Array. *Journal of the Audio Engineering Society*, *65*(7/8), 589–599. https://doi.org/10.17743/jaes.2017.0018

Hendrickx, E., Stitt, P., Messonnier, J.-C., Lyzwa, J.-M., Katz, B. F. G., & Boishéraud, C. de. (2017). Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis. *The Journal of the Acoustical Society of America*, *141*(3), 2011–2023. https://doi.org/10.1121/1.4978612

Higgins, N. C., McLaughlin, S. A., Rinne, T., & Stecker, G. C. (2017). Evidence for cue-independent spatial representation in the human auditory cortex during active listening. *Proceedings of the National Academy of Sciences*, *114*(36), E7602–E7611. https://doi.org/10.1073/pnas.1707522114

Kates, J. M., Arehart, K. H., Muralimanohar, R. K., & Sommerfeldt, K. (2018). Externalization of remote microphone signals using a structural binaural model of the head and pinna. *The Journal of the Acoustical Society of America*, *143*(5), 2666–2677. https://doi.org/10.1121/1.5032326

Katz, B. F. G., & Noisternig, M. (2014). A comparative study of interaural time delay estimation methods. *The Journal of the Acoustical Society of America*, *135*(6), 3530–3540. https://doi.org/10.1121/1.4875714

Klein, F., Werner, S., & Mayenfels, T. (2017). Influences of Training on Externalization of Binaural Synthesis in Situations of Room Divergence. *Journal of the Audio Engineering Society*, *65*(3), 178–187.

Leclère, T., Lavandier, M., & Perrin, F. (2019). On the externalization of sound sources with headphones without reference to a real source. *The Journal of the Acoustical Society of America*, *146*(4), 2309–2320. https://doi.org/10.1121/1.5128325

Li, S., Schlieper, R., & Peissig, J. (2018). The effect of variation of reverberation parameters in contralateral versus ipsilateral ear signals on perceived externalization of a lateral sound source in a listening room. *The Journal of the Acoustical Society of America*, *144*(2), 966–980. https://doi.org/10.1121/1.5051632

Li, S., Schlieper, R., & Peissig, J. (2019). The Role of Reverberation and Magnitude Spectra of Direct Parts in Contralateral and Ipsilateral Ear Signals on Perceived Externalization. *Applied Sciences*, *9*(3), 460. https://doi.org/10.3390/app9030460

Loomis, J. M. (1992). Distal attribution and presence. *Presence: Teleoperators and Virtual Environments*, *1*(1), 113–119.

Lyon, R. F. (1997). All pole models of auditory filtering. In E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, & E. Hecht-Poinar (Eds.), *Diversity in auditory mechanics* (pp. 205–211). World Scientific Publishing, Singapore.

Macpherson, E. A., & Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America*, *111*(5), 2219–2236. https://doi.org/10.1121/1.1471898

Macpherson, E. A., & Middlebrooks, J. C. (2003). Vertical-plane sound localization probed with ripple-spectrum noise. *J. Acoust. Soc. Am.*, *114*(1), 430–445. https://doi.org/10.1121/1.1582174

Macpherson, E. A., & Sabin, A. T. (2007). Binaural weighting of monaural spectral cues for sound localization. *J. Acoust. Soc. Am.*, *121*(6), 3677–3688. https://doi.org/10.1121/1.2722048

Majdak, P., Baumgartner, R., & Jenny, C. (2020). Formation of three-dimensional auditory space. In J. Blauert & J. Braasch, *The Technology of Binaural Understanding*. Springer, Berlin, Heidelberg.

Majdak, P., Baumgartner, R., & Laback, B. (2014). Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization. *Frontiers in Psychology*, *5*(319), 1–10. https://doi.org/10.3389/fpsyg.2014.00319

Majdak, P., & Co. (2020). *The Auditory Modeling Toolbox*. http://amtoolbox.sourceforge.net

May, B. J. (2000). Role of the dorsal cochlear nucleus in the sound localization behavior of cats. *Hearing Research*, *148*(1–2), 74–87. https://doi.org/10.1016/S0378-5955(00)00142-8

Mendonça, C., Campos, G., Dias, P., & Santos, J. A. (2013). Learning Auditory Space: Generalization and Long-Term Effects. *PLOS ONE*, *8*(10), e77900. https://doi.org/10.1371/journal.pone.0077900

Mills, A. W. (1958). On the Minimum Audible Angle. *The Journal of the Acoustical Society of America*, *30*(4), 237–246. https://doi.org/10.1121/1.1909553

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, *40*(10), 1227–1268. https://doi.org/10.1016/S0042-6989(99)00244-8

Reiss, L. A. J., & Young, E. D. (2005). Spectral edge sensitivity in neural circuits of the dorsal cochlear nucleus. *Journal of Neuroscience*, *25*(14), 3680–3691. https://doi.org/10.1523/JNEUROSCI.4963-04.2005

Salminen, N. H., Takanen, M., Santala, O., Lamminsalo, J., Altoè, A., & Pulkki, V. (2015). Integrated processing of spatial cues in human auditory cortex. *Hearing Research*, *327*, 143–152. https://doi.org/10.1016/j.heares.2015.06.006

Schröger, E. (1996). Interaural time and level differences: Integrated or separated processing? *Hearing Research*, *96*(1–2), 191–198. https://doi.org/10.1016/0378-5955(96)00066-4

Shinn-Cunningham, B. G., Santarelli, S., & Kopco, N. (2000). Tori of confusion: Binaural localization cues for sources within reach of a listener. *J Acoust Soc Am*, *107*(3), 1627–1636.

Slattery III, W. H., & Middlebrooks, J. C. (1994). Monaural sound localization: Acute versus chronic unilateral impairment. *Hearing Research*, *75*(1–2), 38–46. https://doi.org/10.1016/0378-5955(94)90053-1

Søndergaard, P., & Majdak, P. (2013). The Auditory Modeling Toolbox. In J. Blauert (Ed.), *The Technology of Binaural Listening* (pp. 33–56). Springer. http://dx.doi.org/10.1007/978-3-642-37762-4_2

Song, Y.-H., Kim, J.-H., Jeong, H.-W., Choi, I., Jeong, D., Kim, K., & Lee, S.-H. (2017). A Neural Circuit for Auditory Dominance over Visual Perception. *Neuron*, *93*(4), 940-954.e6. https://doi.org/10.1016/j.neuron.2017.01.006

Spagnol, S. (2015). On distance dependence of pinna spectral patterns in head-related transfer functions. *The Journal of the Acoustical Society of America*, *137*(1), EL58–EL64. https://doi.org/10.1121/1.4903919

Strutt, J. W. L. R. (1907). On our perception of sound direction. *Philos. Mag.*, *13*, 214–232.

Tardif, E., Murray, M. M., Meylan, R., Spierer, L., & Clarke, S. (2006). The spatio-temporal brain dynamics of processing and integrating sound localization cues in humans. *Brain Research*, *1092*(1), 161–176. https://doi.org/10.1016/j.brainres.2006.03.095

Thornton, M. A., Weaverdyck, M. E., & Tamir, D. I. (2019). The brain represents people as the mental states they habitually experience. *Nature Communications*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-10309-7

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, *6*(4), 391–398. https://doi.org/10.1038/nn1032

van der Heijden, K., Rauschecker, J. P., de Gelder, B., & Formisano, E. (2019). Cortical mechanisms of spatial hearing. *Nature Reviews Neuroscience*, *20*(10), 609–623. https://doi.org/10.1038/s41583-019-0206-5

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*(1), 242–268. https://doi.org/10.1111/nyas.14321

Weber, E. H. (1848). On the circumstances under which one is led to refer sensations to external objects. *Proceedings of the Royal Saxon Society for Science in Leipzig*, *2*, 226–237.

Wood, K. C., Town, S. M., & Bizley, J. K. (2019). Neurons in primary auditory cortex represent sound source location in a cue-invariant manner. *Nature Communications*, *10*(1), 1–15. https://doi.org/10.1038/s41467-019-10868-9

Zhang, P. X., & Hartmann, W. M. (2010). On the ability of human listeners to distinguish between front and back. *Hear Res*, *260*(1–2), 30–46.