Acta Acustica 2020, **4**, 21
© S. Li et al., Published by EDP Sciences, 2020
https://doi.org/10.1051/aacus/2020020

Available online at:
https://acta-acustica.edpsciences.org

Scientific Article Open Access

Modeling perceived externalization of a static, lateral sound image

Song Li[1,*], Robert Baumgartner[2], and Jürgen Peissig[1]

[1] Institute of Communications Technology, Gottfried Wilhelm Leibniz Universität Hannover, 30167 Hannover, Germany
[2] Acoustics Research Institute, Austrian Academy of Sciences, 1040 Vienna, Austria

Received 20 April 2020, Accepted 23 September 2020

Abstract – Perceived externalization is a relevant feature to create an immersive acoustic environment with headphone reproduction. In the present study, listener-specific acoustic transfer characteristics for an azimuth angle of 90° were modified to investigate the role of monaural spectral cues, interaural level differences (ILDs), and temporal fluctuations of ILDs on perceived externalization in anechoic and reverberant environments. Listeners' ratings suggested that each acoustic cue was important for perceived externalization. If only one correct acoustic cue remained in the ear signals, the sound image could not be perceived as fully externalized. Reverberation did reduce but not eliminate the influences of monaural spectral and ILD cues on perceived externalization. Additionally, the spectral details of the ipsilateral ear signal were more important for perceived externalization than those in the contralateral ear signal. A computational model was proposed to quantify those relationships and predict externalization ratings by comparing the acoustic cues extracted from the target (modified) and template (non-processed) binaural signals after several auditory processing steps. The accuracy of predicted externalization ratings was higher than 90% under all experimental conditions.

Keywords: Externalization, Monaural spectral cues, Interaural level differences, ILD temporal fluctuations

1 Introduction

Sound externalization describes the ability to attribute auditory signals to external sources and perceive them located at some distance [1]. Thanks to binaural technology, it is possible to reproduce externalized virtual sound images over headphones [2]. A standard method to generate binaural sounds is by convolving an anechoic audio signal with a pair of head-related impulse responses (HRIRs) [3], or binaural room impulse responses (BRIRs) [4]. When the spatial properties of virtual sound images do not match those of the listener's natural acoustic exposure, the externalization can be easily distorted, leading to sound images being perceived inside the head or close to the skull. Perceived externalization is a matter of degree and considered to mediate the perception of distance [5, 6]. Consequently, changes in perceived distance can indicate changes in the ability to externalize the percept and are most commonly used for its assessment. Over the years, various psychoacoustic experiments have been conducted to investigate relevant cues contained in head-related transfer functions or binaural room transfer functions (HRTFs or BRTFs, frequency domain representation of HRIRs or BRIRs) that are necessary for creating externalized virtual sound images, but their combined influences are still poorly understood.

In free-field conditions, synthesized virtual sound sources could be perceived as well-externalized using individually measured HRTFs. Hartmann and Wittenberg [6] revealed that interaural time differences (ITDs)/interaural phase differences (IPDs) at low frequencies and interaural level differences (ILDs) at all frequencies were important for perceived externalization. However, the virtual sound could not be perceived as well-externalized if only the correct ILDs were preserved, but the spectral information in the ear signals was distorted. The listening test performed by Kulkarni and Colburn [7] reported that smoothing the magnitude spectra of HRTFs while scrambling intensities within third octave bands affected only the perception of elevation but not externalization. In contrast, Baumgartner et al. [8] demonstrated both behaviorally and neurally that spectral flattening of HRTFs reduced the degree of perceived externalization. In both cases, the manipulation of the magnitude spectra of HRTFs caused not only a change of the monaural but also the interaural spectral information contained in HRTFs.

In reverberant conditions, the room acoustics also affect perceived externalization [5, 9–11]. Several studies have shown that the reverberation between 20 ms and 80 ms, i.e., the early reflection part, had an influence on perceived externalization; increasing the reverberant part to durations longer than about 80 ms did not change perceived externalization further [12–15]. Werner et al. [10] demonstrated that

[*] Corresponding author: song.li@ikt.uni-hannover.de

the degree of perceived externalization was reduced when the acoustics of the synthesized room differed from those of the playback room. Catic et al. [14] studied the role of monaural and binaural reverberation on perceived externalization. Their results pointed out that monaural reverberation was sufficient to externalize lateral sound images, while binaural reverberation was necessary for the externalization of frontal sound sources. In addition, the perceptual data were compared with the monaural direct-to-reverberant energy ratio (DRR), and binaural cues, such as ILD temporal fluctuations, interaural coherence (IC) and IC temporal fluctuations. The change of externalization ratings was highly correlated with all three reverberation-related binaural cues over various experimental conditions in their study. In contrast, listeners' judgments were not well reflected in the change in DRR. This means that the reverberation-related binaural cues could be utilized as indicators to predict perceived externalization. Furthermore, Li et al. [15] separately manipulated the reverberation heard by each ear to investigate the relative influence of reverberation in contralateral versus ipsilateral ear signals on the externalization of a lateral sound source. The reverberation at the contralateral ear was more important to externalization than that at the ipsilateral ear. Additionally, various reverberation-related binaural cues were contrasted in that study as being part of a quantitative model for predicting the externalization results. The model based on ILD temporal fluctuations showed the best performance for their experiments.

Reverberation reduces the relevance of spectral details [16] but spectral details in the direct path component of BRTFs remain important [17]. In contrast, the spectral information contained in the reverberation part did not noticeably affect the degree of externalization [17]. Li et al. [18] smoothed the spectral magnitude of direct sounds in each ear separately to investigate the role of spectral information of direct parts in contralateral and ipsilateral ear signals on perceived externalization. The results showed that the spectral information in direct parts of the ipsilateral ear had more influence on externalization than that of the contralateral ear. It remains unclear to what degree the experimental results in previous studies can be attributed to the degradation of spectral details, the ILD deviations, or the combination of both cues. Hassager et al. [17] proposed an ILD-based externalization model, which was appropriate to explain the average experimental results obtained in their study. However, they did not consider other acoustic cues related to perceived externalization even though it is already known that the correct ILD information alone is not sufficient to well externalize virtual sound images [6].

In most previous studies, monaural spectral cues, ILDs and temporal fluctuations of ILDs were separately investigated. The present study aimed at a quantitative model explaining the interplay of these important acoustic cues in perceived externalization by following a template-matching procedure [17, 19]. The proposed model extends previous approaches [15, 17, 19] by incorporating all three relevant acoustic cues to jointly predict the degree of externalization. The perceptual weights of the different acoustic cues and the binaural weighting between ipsi- and contralaterally processed cues were derived from different externalization experiments under static listening conditions. To this end, the model considers the reduced relevance of spectral details for perceived externalization with increasing amount of reverberation. Furthermore, we investigated frequency dependence of ILDs for predicting externalization results.

This paper is organized as follows. Section 2 describes the proposed externalization model. To address open issues in the design of this model and to parameterize it based on listener-specific data we have designed a series of experiments as described in Section 3. The calculation of the weighting factor for each model parameter and the predicted externalization results are presented in Section 4. The experimental results, as well as model components and limitations are discussed in Section 5. Finally, conclusions and directions for future work are summarized in Section 6.

2 Externalization model

2.1 Concept and overview

Plenge [20] introduced a conceptual localization model to explain the perception of externalization by humans, consisting of a long-term and a short-term memory. In this model, localization cues represented in HRIRs are stored in the long-term memory. An adaptation to altered head-related auditory localization cues can take several days [21]. The short-term memory represents reverberation-related acoustic cues. In contrast to the information stored in the long-term memory, adaptation of the short-term memory is required every time the listening environment changes. A virtual sound source is expected to be perceived as well-externalized if the target binaural sounds provide "information" similar to that stored in both memories.

Figure 1 shows the structure of our proposed externalization model, which follows Plenge's conceptual framework. In this model, the magnitude spectra and ILDs are represented as the long-term memory information, and the ILD temporal fluctuations are used as an indicator of the reverberation-related acoustic cues stored in the short-term memory. In order to compare the information stored in the long-term memory, the direct sound component should be first extracted from the binaural signals. In this study, the direct sound part ("target" and "template") is simulated by convolving the direct part extracted from the BRIR and the input signal (white Gaussian noise). Next, the obtained direct sound part is filtered through a gammatone filter bank [22] with a spacing and bandwidth of one equivalent rectangular bandwidth (ERB) [23], which is a common approximation of the cochlear filtering process. It is used to derive temporally long- or short-term averaged "excitation patterns", where the filtered signal in each channel is temporally averaged in terms of root mean square (RMS). After that, the spectral gradients (SGs) [19, 24]

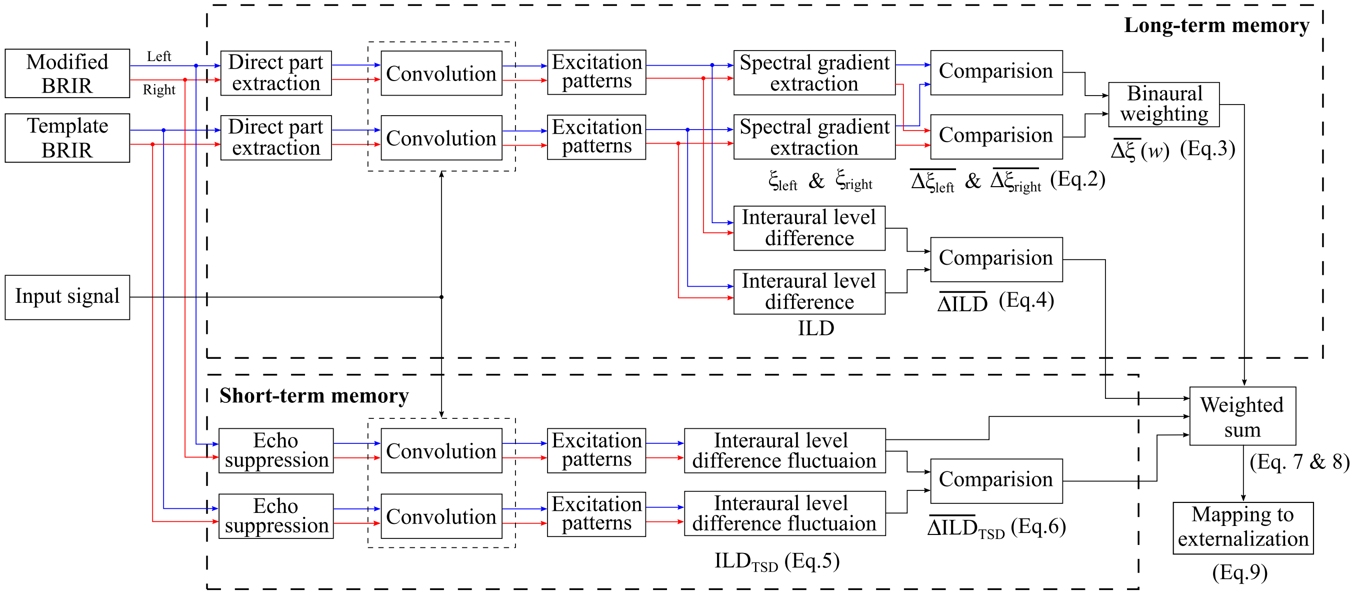
Long-term memory
Modified BRIR
Left
Right
Template BRIR
Input signal
Direct part extraction
Direct part extraction
Convolution
Convolution
Excitation patterns
Excitation patterns
Spectral gradient extraction
Spectral gradient extraction
ξ_{left} & ξ_{right}
Comparision
Comparision
$\overline{\Delta\xi_{\text{left}}}$ & $\overline{\Delta\xi_{\text{right}}}$ (Eq.2)
Binaural weighting
$\overline{\Delta\xi}(w)$ (Eq.3)
Interaural level difference
Interaural level difference
ILD
Comparision
$\overline{\Delta\text{ILD}}$ (Eq.4)
Short-term memory
Echo suppression
Echo suppression
Convolution
Convolution
Excitation patterns
Excitation patterns
Interaural level difference fluctuaion
Interaural level difference fluctuaion
ILD_{TSD} (Eq.5)
Comparision
$\overline{\Delta\text{ILD}}_{\text{TSD}}$ (Eq.6)
Weighted sum
(Eq. 7 & 8)
Mapping to externalization
(Eq.9)

Figure 1. Structure of the proposed externalization model, consisting of a short-term and a long-term memory. In the long-term memory, the spectral gradients and ILDs are extracted from the direct sound part of the target signal in each frequency channel of a gammatone filter bank. In the short-term memory, the ILD temporal standard deviations are obtained from the echo-suppressed reverberant signals in each frequency channel. The deviations of these three acoustic cues from the template signals are summed up with different weighting factors and mapped to externalization ratings.

and ILDs [17] of the target signal are calculated for each frequency band and compared with those stored in the long-term memory ("template"). The deviations are weighted based on the local room acoustics to represent the relevance of the SGs and ILDs on perceived externalization with different listening environments. To compare the information stored in the short-term memory, the ILD temporal fluctuations are obtained from the echo-suppressed reverberant signals in each frequency band and compared with those stored in the short-term memory. The deviations of these acoustic cues from the template signals are summed up with different weighting factors and mapped to externalization ratings.

2.2 Processing stages

2.2.1 SG comparison

To quantify the spectral information of the sound source in each ear, the SGs, defined as the excitation differences (in dB) between neighboring pairs of frequency channels, $\xi_k(i)$, are calculated as [19]:

$$\xi_k(i) = M_k(f_{c,i}) - M_k(f_{c,i-1}), \quad \text{for} \quad i = 2, 3, ..., N, \tag{1}$$

where the index k represents the side of the ear, i.e., $k \in \{\text{left}, \text{right}\}$. $M_k(f_{c,i})$ denotes the excitation in a single frequency band centered at $f_{c,i}$ (center frequency of the ith frequency channel), and N denotes the number of frequency channels allocated from 0.2 to 16 kHz. Spectral shapes are compared based on the absolute difference of the SGs between the "target" (convolution of the stimulus with modified HRTFs/BRTFs) and the "template" (convolution of the stimulus with original HRTFs/BRTFs), and the mean deviation over frequencies is normalized by dividing through the averaged SG from the template signals:

$$\overline{\Delta\xi_k} = \frac{\sum_{i=2}^{N} |\xi_{k,\text{target}}(i) - \xi_{k,\text{template}}(i)|}{\sum_{i=2}^{N} |\xi_{k,\text{template}}(i)|}. \tag{2}$$

The normalization process is not performed within frequency channels, since the zero gradient point is rather arbitrary from a neural perspective and we do not want to introduce any biases through normalization. The SG-based comparison metrics calculated for each ear individually are weighted by a binaural weighting factor, w, determining the contribution of the spectral information of the left and right ear to perceived externalization:

$$\overline{\Delta\xi}(w) = w\overline{\Delta\xi_{\text{left}}} + (1 - w)\overline{\Delta\xi_{\text{right}}}, \tag{3}$$

where w is limited between zero and one.

2.2.2 ILD comparison

Because ILDs are naturally larger at high frequencies, the same absolute ILD offset leads to a smaller relative change in ILD when applied to a high as compared to a low frequency sound. Thus, in the comparison stage, as opposed to the normalization for SG deviations (cf. Eq. (2)), normalized ILD deviations are calculated within each frequency band and averaged across frequency bands according to [17]:

$$\overline{\Delta \mathrm{ILD}} = \frac{1}{N} \sum_{i=1}^{N} \frac{|\mathrm{ILD}_{\mathrm{target}}(f_{c,i}) - \mathrm{ILD}_{\mathrm{template}}(f_{c,i})|}{|\mathrm{ILD}_{\mathrm{template}}(f_{c,i})|}, \quad (4)$$

where $\mathrm{ILD}_{\mathrm{target}}(f_{c,i})$ and $\mathrm{ILD}_{\mathrm{template}}(f_{c,i})$ denote the ILD of target and template signals centered at $f_{c,i}$, respectively.

2.2.3 Temporal fluctuation comparison

Li et al. [15] showed that ILD temporal fluctuations corresponded well to externalization ratings in response to modified reverberation level. Thus, the ILD temporal fluctuations of the target binaural signals are extracted here to compare the information stored in the short-term memory (“template”). The target binaural signals are first processed through an echo-suppression mechanism motivated by the “precedence effect” [25, 26], which is achieved by multiplying the BRIR with a time window that had a value of one up to 2.5 ms (duration of the direct part), followed by zeros up to 10 ms (echo-suppression) and a transition from zero to one using a raised-cosine window from 10 ms to 15 ms (for all measured BRIRs in this study) [14, 15]. Catic et al. [14] demonstrated that this approximation of the “precedence effect” is important to predict perceived externalization. Then, the echo-suppressed binaural signals are filtered through a gammatone filter bank with a bandwidth of one ERB. Afterwards, the short-term values of ILDs are calculated in a 20 ms long Hann window with a 50% overlap over the duration of binaural signals (99 windowed frames for a 1 s long signal) in each frequency band. The ILD temporal fluctuations are defined as the standard deviation across the short-term ILDs. In each frequency band centered at f_c, the ILD temporal standard deviation (ILD TSD) is calculated as:

$$\mathrm{ILD}_{\mathrm{TSD}}(f_c) = \sqrt{\frac{1}{N_{\mathrm{frame}} - 1} \sum_{n=1}^{N_{\mathrm{frame}}} \left(\mathrm{ILD}(f_c, n) - \overline{\mathrm{ILD}(f_c)}\right)^2}, \quad (5)$$

where N_{frame} is the number of frames contained in the binaural signals. $\mathrm{ILD}(f_c, n)$ and $\overline{\mathrm{ILD}(f_c)}$ represent the ILD in the nth frame and the averaged ILD across the time frames, respectively.

In the present model, ILD TSDs are considered to represent reverberation-related cues. However, also in anechoic conditions, random fluctuations of the source signal cause the ILD TSDs to be larger than zero. Hence, if put in relation to the small values of template ILD TSDs, minor changes in ILD TSDs caused by HRTF manipulations would have the potential to lead to large unwanted deviations. Instead, absolute ILD TSD deviations are calculated in the comparison stage, and an arbitrary scaling factor is introduced for consistency reasons of unitless deviation metrics. Hence, the absolute deviation is divided through the averaged ILD TSDs over frequencies of a reference acoustic environment, $\overline{\mathrm{ILD}}_{\mathrm{TSD,reference}}$. Here, the listening room (cf. Experiment E) is considered as the reference acoustic environment:

$$\overline{\Delta \mathrm{ILD}}_{\mathrm{TSD}} = \frac{\sum_{i=1}^{N} \left|\mathrm{ILD}_{\mathrm{TSD,target}}(f_{c,i}) - \mathrm{ILD}_{\mathrm{TSD,template}}(f_{c,i})\right|}{\overline{\mathrm{ILD}}_{\mathrm{TSD,reference}}}, \quad (6)$$

where $\mathrm{ILD}_{\mathrm{TSD,target}}(f_{c,i})$ and $\mathrm{ILD}_{\mathrm{TSD,template}}(f_{c,i})$ denote the ILD TSDs of target and template signals centered at $f_{c,i}$, respectively.

2.2.4 Reverberation-related weighting

The influences of SGs and ILDs on perceived externalization are mainly studied in anechoic conditions. However, the role of SGs and ILDs on externalization may be reduced in the reverberant environment compared to that in the anechoic environment, i.e., the influence of SGs and ILDs decreases with the increasing reverberation. To represent this effect, short- and long-term memory evaluation results are weighted by a factor γ depending on the presented amount of reverberation-related features:

$$\gamma = 1 - b_{\gamma} \frac{\overline{\mathrm{ILD}}_{\mathrm{TSD,template}}}{\overline{\mathrm{ILD}}_{\mathrm{TSD,reference}}}, \quad (7)$$

where γ is limited between zero and one and b_{γ} is a weighting factor. $\overline{\mathrm{ILD}}_{\mathrm{TSD,template}}$ denotes the averaged ILD TSDs over frequencies of the template signal, representing the current acoustic environment. $\overline{\mathrm{ILD}}_{\mathrm{TSD,reference}}$ represents the averaged ILD TSDs over frequencies in a reference acoustic environment as introduced in Equation (6). This reduction term is applied for reducing the influences of SGs and ILDs based on the current acoustic environment, i.e., γ is close to one for anechoic and smaller than one for our most reverberant condition.

2.2.5 Mapping to externalization

After the comparison stage in both memories, the differences of SGs, ILDs, and ILD TSDs between the “target” and the “template” are summed up with different weighting factors, which can be expressed as [27]:

$$\Delta m = \gamma\left(b_{\mathrm{ILD}} \overline{\Delta \mathrm{ILD}} + b_{\xi} \overline{\Delta \xi}(w)\right) + b_{\mathrm{ILDTSD}} \overline{\Delta \mathrm{ILD}}_{\mathrm{TSD}}, \quad (8)$$

where b_{ILD}, b_{ξ} and $b_{\mathrm{ILD\ TSD}}$ are weighting factors for deviations of acoustic cues, respectively. The mapping between the objective measures and the externalization ratings is represented by an exponential function [15, 17]:

$$E = a e^{-\Delta m} + c, \quad (9)$$

where a and c are two mapping parameters. In Section 3, various psychoacoustic experiments are designed to determine the weighting parameters for the model components. The derivation of the weighting parameters is described in Section 4.

3 Experiments

3.1 General methods

3.1.1 Measurement of individual impulse responses

Individual HRIRs and BRIRs were measured in an anechoic room (4.7 m × 4.3 m × 3.5 m, IAC Acoustics, lower frequency boundary around 200 Hz) and a listening room (6.7 m × 4.8 m × 3.2 m, reverberation time about 260 ms) both located at the Institute of Communications Technology (IKT) of Leibniz Universität Hannover. A loudspeaker (Neumann KH 120 A) was placed at an azimuth angle of 90° (left side) with a distance of 1.5 m from the subject, and served as a sound source for the impulse response recording. A pair of commercial binaural microphones (Madness MM-BSM-8) was placed at the entrance of each subject's ear canals for the measurement. The HRIRs and BRIRs were measured using a 5 s-long exponential sweep [28], with 10 repetitions and a sampling frequency of 44.1 kHz. In the case of the HRIR recording, the impulse response measured was truncated by a 2.5 ms long time window (after the propagation time from the loudspeaker to the listener, i.e., onset delay) with a 0.5 ms long half raised-cosine fall time. The HRIRs measured were equalized by a reference measurement with the in-ear microphones placed at the location of the center of the subject's head without the subject being present [29]. For the BRIR measurement, the impulse responses measured were truncated by a 260 ms long time window (after the onset delay) with a 10 ms long half raised-cosine fall time [15].

3.1.2 Experimental paradigm

Five subjects (one female and four male) with normal hearing and aged between 24 and 30 participated in the experiments. Three of them (including the first author) had participated in similar experiments before.

In a series of five experiments, we investigated how the degree of perceived externalization was affected by (A) changes in ILDs, (B) changes in spectral information while maintaining original ILDs, (C) interaural reductions in spectral details, (D) deviations in ILDs and spectral details, and (E) deviations in ILDs, spectral information, and reverberation. These experiments were designed to determine the model weightings of monaural spectral cues, ILDs, and the ILD temporal fluctuations concerning perceived externalization while trying to isolate individual cues. Additionally, some experiments/experimental conditions were only used to evaluate the performance of the model. The experiments focusing on the influence of the spectral information and ILDs on perceived externalization were conducted in an anechoic chamber, whereas the experiment about the role of reverberation on externalization was performed in a listening room. In this study, the HRIRs and the direct parts of the BRIRs were represented as minimum-phase components, followed by all-pass filters [17, 30]. In so doing, the magnitude spectra of the minimum-phase components could be manipulated while preserving the phase information. The stimuli used in the experiments were generated by the convolution of a 1-s long white Gaussian noise (200 Hz–16 kHz) with modified HRIRs/BRIRs (the same noise signal was used for generating different stimuli). The audio signals were presented via headphones at a sound pressure level (SPL) of about 67 dB.

Each listener sat in a chair, listened to the stimuli under test with a pair of individually compensated Sennheiser HD800 headphones according to Schärer and Lindau [31]. The headphone transfer function (HpTF) was measured for each listener with 10 repetitions (headphones repositioned), and the corresponding compensation filter was obtained by applying the least-squares inversion method in combination with a frequency-dependent regularization to the averaged HpTF [31]. To create smooth transitions from the in-head localized sound source to the fully externalized sound source (at the position of the real sound source), a distance-related measure is commonly applied for externalization experiments. In this study, subjects were asked to assess the degree of perceived externalization using a subjective rating scale similar to that used in our previous study [15], as shown in Table 1. They were asked to rate each stimulus using a slider with a step-size of 0.1 between 0 and 3 and to ignore audible artifacts that do not affect their externalization perception. In addition, listeners were not allowed to turn their heads during each experiment.

Before each listening experiment, each subject was asked to listen to the stimuli once in order to become familiar with each stimulus presented in the experiment. Each experiment was tested four times (the stimuli were in random order), and the externalization score was taken as the mean of these four scores. The loudspeaker was always presented during each experiment. As anchors, subjects listened to the original stimulus played back through the loudspeaker and were informed that such stimulus should act as a "fully-externalized" sound (externalization rating = 3). The subjects had to put off the headphone for listening to the reference anchor signals played back by the loudspeaker. Additionally, diotic playback of the source signal acted as a "fully-internalized" anchor sound (externalization rating = 0). Subjects could listen to the anchor sounds at any time during each experiment.

The present study focused on a virtual sound source with an incidence angle of 90°. This extreme lateralization allows to determine the range of the binaural weighting factor applied to monaural information across different azimuths. The weighting factor for a frontal sound source is expected to be 0.5, and for other azimuth angles, it can be interpolated [24]. In addition, only a 90° sound source allows to manipulate broadband ILDs without affecting perceived lateralization (see Experiment A). Further, the externalization of a lateral sound source is hardly affected by potential head movements we were not able to control with our setup [32].

3.2 Experiment A: influence of ILDs

Experiment A aimed to investigate the influence of ILDs on perceived externalization, and further to obtain the weighting factor of the ILD cue while maximally isolating

Table 1. Subjective rating scale for perceived externalization.

Degree	Meaning of the degree
3	The sound is externalized and at the position of the loudspeaker.
2	The sound is externalized but not as far as the loudspeaker.
1	The sound is not well-externalized. It is at my ear.
0	The sound is in my head.

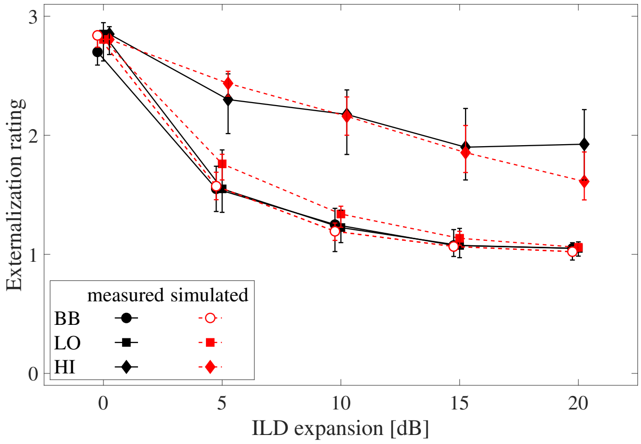

Figure 2. Median values of externalization ratings (solid lines) and model simulations (dashed lines, open and filled symbols for mapped and predicted results, respectively) with 95% confidence intervals (CIs) for ILD expansions in three different frequency ranges ("BB", "LO" and "HI").

spectral changes. It is not possible to manipulate the frequency-dependent ILD information, i.e., the ILD contrast over frequencies, without changing the spectral information of the HRTF at each ear. A feasible way to manipulate the ILD information without changing the magnitude spectra in HRTFs was to independently control the level of the sounds delivered to each ear, i.e., manipulating only the broadband ILD [33]. By attenuating the sound level at the ipsilateral ear (left ear), i.e., reducing the ILD magnitude, the sound source might be perceived at a more central lateral angle, or subjects could even hear two split sound images due to the contradictory ITD and ILD information. Hence, we expanded the ILD by attenuating the signal delivered to the contralateral ear (right ear) in order to maintain perceived lateralization of the sound image at 90°. In Experiment A, the sound level at the right (contralateral) ear was attenuated by 0, 5, 10, 15, and 20 dB at the broadband (BB: 0.2–16 kHz), only at the low frequency range (LO: 0.2–3 kHz) and the high frequency range (HI: 3–16 kHz), while the SPL of the left (ipsilateral) ear signal remained unchanged. The spectrum of the modified HRTF in the right ear (contralateral ear), $|\mathrm{HRTF}_{\mathrm{right,mod}}(f)|$, can be expressed as:

$$\left|\mathrm{HRTF}_{\mathrm{right,mod}}(f)\right| = \frac{\left|\mathrm{HRTF}_{\mathrm{right}}(f)\right|}{10^{\frac{A}{20}}}, \quad \text{for} \quad f \in \{\mathrm{BB}, \mathrm{LO}, \mathrm{HI}\}, \tag{10}$$

where A denotes different attenuations in dB. This experiment was divided into three blocks based on the three manipulated frequency ranges. Five stimuli with different attenuations were presented in each section. A total of 60 stimuli (including repetitions) were assessed by each listener.

Figure 2 shows the externalization ratings for ILD expansions in the three frequency ranges. For comparison, the simulation results calculated according to our proposed externalization model (refer to Sect. 4) are also plotted. Overall, the degree of externalization ratings decreased with increasing ILD difference in all three frequency ranges. Both, the ILD expansions (Friedman test: $\chi^2(4) = 55.6$, $p = 0.01$) and the manipulated frequency ranges (Friedman test: $\chi^2(2) = 32.1$, $p = 0.01$) had significant effects on perceived externalization. The externalization ratings were noticeably reduced by an ILD increase of 5 dB at the low and the whole frequency ranges. The sound image was perceived as being at the ear (externalization ratings ≈ 1) by further increasing the ILD, and there were no significant differences in externalization ratings between the low and the whole frequency range being modified (Friedman test: $\chi^2(1) = 0.89$, $p = 0.4$). In contrast, the decrease in externalization ratings was smaller when the ILD at high frequencies increased. The sound source was still perceived as externalized even with an increase of 20 dB ILD at high frequencies.

3.3 Experiment B: influence of spectral details with unchanged ILDs

Experiment B aimed to investigate the influence of the spectral information contained in HRTFs on perceived externalization. The spectral ILD remained unchanged, while the magnitude spectrum of the HRTF in the ipsilateral ear (left ear) was smoothed by using a 4th-order gammatone filter, $|H(f, f_c)|$, at the center frequency of f_c with the bandwidth of $b(f_c)$ [17, 18]:

$$\left|\mathrm{HRTF}_{\mathrm{left,mod}}(f_c)\right| = \sqrt{\frac{\int_0^\infty \left|\mathrm{HRTF}_{\mathrm{left}}(f)\right|^2 \left|H(f, f_c)\right|^2 \mathrm{d}f}{\int_0^\infty \left|H(f, f_c)\right|^2 \mathrm{d}f}}. \tag{11}$$

The magnitude spectrum of the 4th-order gammatone filter was approximated as [34]:

$$|H(f, f_c)| = \left|\left(\frac{b(f_c)}{\mathrm{j}(f - f_c) + b(f_c)}\right)^4\right| \quad \text{with} \quad j = \sqrt{-1}. \tag{12}$$

The smoothing level of the magnitude spectrum was achieved by controlling the bandwidth $b(f_c)$ of the gammatone filter [6]:

$$b(f_c) = B \times \frac{24.7\,(0.00437 \times f_c)}{2\sqrt{2^{1/4} - 1}} = 0.1241B \times f_c, \tag{13}$$

where B represents the bandwidth factor relative to a value of one ERB. In this experiment, the spectral

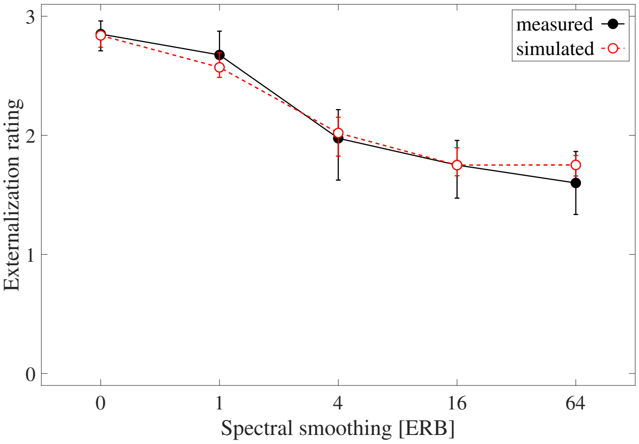

Figure 3. Median values of externalization ratings and mapped results for smoothed spectral magnitude in the HRTF of the ipsilateral ear while maintaining the original ILD. All other conventions are as in Figure 2.

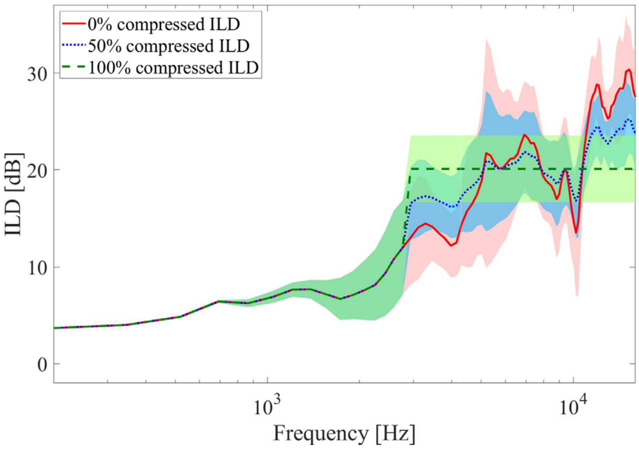

Figure 4. Averaged ILD across subjects with 0% (red solid line), 50% (blue dotted line) and 100% (green dashed line) compression factors. Shaded areas denote 95% CIs of means.

magnitude was smoothed with a bandwidth factor $B \in \{0, 1, 4, 16, 64\}$. $B = 0$ denotes the unprocessed spectral magnitude. The magnitude spectrum of the HRTF in the contralateral ear was adapted to maintain the spectral ILD. In Experiment B, a total of 20 stimuli (including repetitions) were evaluated by each listener.

Figure 3 shows the externalization results by smoothing the spectral details in the HRTF of the ipsilateral ear while maintaining the original ILD. The externalization ratings reduced significantly with bandwidth above one ERB (Friedman test: $\chi^2(4) = 20$, $p = 0.01$). The median value of the externalization rating was about 1.6 for the most severe smoothing. This means that maintaining the correct spectral ILD information might be sufficient to externalize a lateral sound source, but not enough to externalize it well.

3.4 Experiment C: influence of interaural spectral details

Experiment C was designed to study the influence of interaural (as opposed to monaural) spectral details in HRTFs. However, one can not change the interaural spectral contrast while maintaining the original magnitude spectra in both ears. The spectral information in HRTFs is more pronounced at high frequencies than at low frequencies due to the reflections and diffractions caused by the pinnae. Therefore, the spectral ILD contrast was compressed at high frequencies with different compression factors, while the magnitude spectrum of the HRTF was preserved in one ear. The ILD between 3 kHz and 16 kHz in dB was modified:

$$\mathrm{ILD}_{\mathrm{mod}}(f) = (1 - C)\,\mathrm{ILD}(f) + C\,\frac{1}{\sum\limits_{k\in f} w(k)} \sum_{k\in f} w(k)\,\mathrm{ILD}(f), \tag{14}$$

where C represents the compression factor, and $w(k)$ is a frequency-dependent weight that approaches the resolution of auditory filters by the across-frequency derivative of ERBs frequencies (cf. [8]). The ILD spectral contrast was compressed using compression factors, C, of 0%, 25%, 50%, 75%, and 100%. For C equal to 0%, the original ILD was presented. In contrast, if C was equal to 100%, the ILD was constant between 3 kHz and 16 kHz. Figure 4 illustrates an example of processed ILDs averaged across subjects with compression factors of 0% (red solid line), 50% (blue dotted line) and 100% (green dashed line). Two conditions were considered in this experiment: (i) The spectral magnitude of the HRTF was changed only at the contralateral ear while reducing the ILD contrast (condition “contra”). (ii) The spectral magnitude of the HRTF was only changed in the ipsilateral ear while the ILD contrast was reduced (condition “ipsi”). Five stimuli with different ILD contrasts were generated by modified HRTFs in each condition and presented over headphones. In total, 40 stimuli (including repetitions) were assessed by each listener.

Figure 5 shows the externalization ratings by reducing the ILD contrast at high frequencies while manipulating the spectral details in the ipsilateral (circles) or the contralateral ear (squares). Both the ILD contrast (Friedman test: $\chi^2(4) = 38.1$, $p = 0.01$) and the side of the modified ear (Friedman test: $\chi^2(1) = 14.4$, $p = 0.01$) had significant effects on externalization results. The externalization ratings decreased substantially for compression factors above 25%. Furthermore, the degree of externalization reduced more for the “ipsi” condition than the “contra” condition, meaning that the spectral information contained in the HRTF of the ipsilateral ear was more important for perceived externalization than that of the contralateral ear.

3.5 Experiment D: influences of ILDs and spectral details

Experiment D was designed to target the additive influences of monaural spectral and ILD cues on perceived externalization. To this end, the same smoothing method as described in Experiment B was applied, but in contrast to

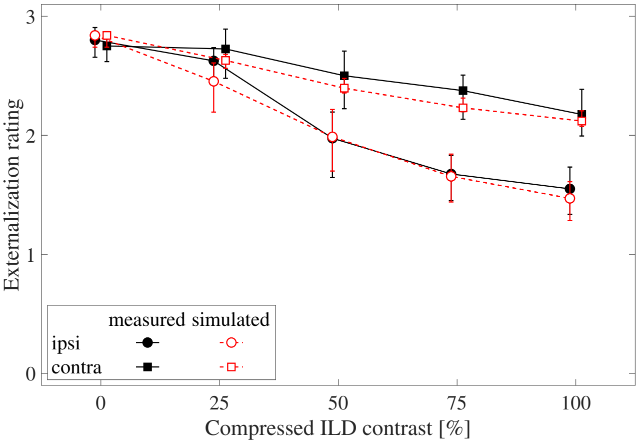

Figure 5. Median values of externalization ratings and the mapped results for reduced ILD contrasts with ipsilateral (condition "ipsi") versus contralateral (condition "contra") spectral distortions. All other conventions are as in Figure 2.

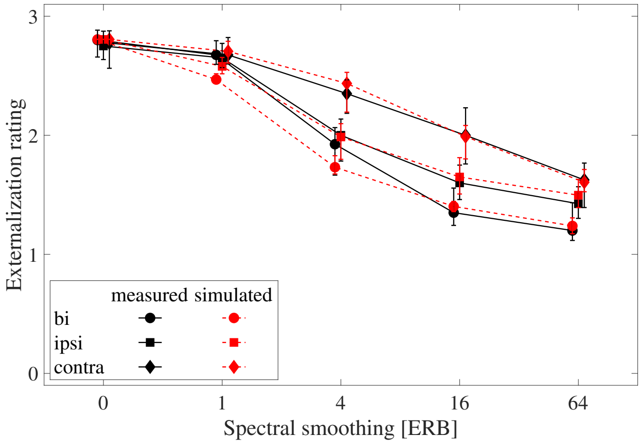

Figure 6. Median values of externalization ratings and predicted results by reducing spectral details in HRTFs of both ears (condition "bi"), the ipsilateral ear (condition "ipsi") or the contralateral ear (condition "contra"). All other conventions are as in Figure 2.

Experiment B, spectral ILDs were not maintained by spectral compensation in the other ear. The magnitude spectra were smoothed by different smoothing factors, $B \in \{0, 1, 4, 16, 64\}$, in HRTFs of (i) both ears (condition "bi"), (ii) the ipsilateral ear (condition "ipsi") and (iii) the contralateral ear (condition "contra"). The five different smoothing factors were applied for each condition, and each listener evaluated a total of 60 audio sequences (including repetitions).

Figure 6 shows the externalization ratings by reducing the spectral details in the HRTF of both ears (circles), the ipsilateral ear (squares), and the contralateral ear (diamonds). Results of statistical tests showed that both the spectral smoothing levels (Friedman test: $\chi^2(4) = 64.6$, $p = 0.01$) and the smoothing conditions (Friedman test: $\chi^2(2) = 15.8$, $p = 0.01$) had significant effects on perceived externalization. The externalization ratings were noticeably reduced with bandwidths above one ERB for all three conditions. For bandwidths larger than one ERB, the degree of externalization decreased dramatically with increasing bandwidths for all three conditions. Furthermore, the ratings for "bi" and "ipsi" conditions reached a low level of about 1.5 earlier than those for the "contra" condition. This result indicated that smoothing the magnitude spectra of HRTFs at the ipsilateral ear was more effective than at the contralateral ear to reduce externalization.

3.6 Experiment E: influences of ILDs, spectral information and reverberation

Experiment E aimed to investigate the interaction among reverberation, ILDs and the monaural spectral cues on perceived externalization. For that, the spectral magnitude of the direct part in the BRIR was smoothed with different smoothing factors, $B \in \{0, 1, 4, 16, 64\}$, applied bilaterally using the same method as in Experiment D. The direct part was extracted with a 2.5 ms long window [14, 15] with a 0.5 ms long half raised-cosine fall time. The reverberant part was obtained by subtracting the direct part from the BRIR. Additionally, the level of the reverberation was reduced by multiplying the reverberant part with a scaling factor α ($0 < \alpha < 1$) in the time domain:

$$\mathrm{BRIR}_{\mathrm{mod}}(t) = \mathrm{BRIR}_{\mathrm{direct,mod}}(t) + (1 - \alpha)\,\mathrm{BRIR}_{\mathrm{reverb}}(t). \tag{15}$$

The modified direct part, $\mathrm{BRIR}_{\mathrm{direct,mod}}(t)$, was calculated according to Equation (11). The amount of reverberation was reduced with scaling factors of 0%, 25%, 50%, 75%, and 100%. $\alpha = 0\%$ stands for the non-processed reverberant part. The reverberation was not present if α was equal to 100%. The BRIRs measured were manipulated with different smoothing and reverberation reduction levels. This experiment was divided into five sections, and in each section, five stimuli generated by modified BRIRs with the same smoothing factor but different reverberation reduction levels were presented. In total, 100 audio sequences (including repetitions) should be evaluated by each listener.

Figure 7 shows the externalization ratings for different spectral smoothing ($B \in \{0, 1, 4, 16, 64\}$) and reverberation reduction (0%, 25%, 50%, 75%, and 100%) levels across subjects. Compressing the reverberant part (Friedman test: $\chi^2(4) = 106.2$, $p = 0.01$) and smoothing the spectral details (Friedman test: $\chi^2(4) = 90.1$, $p = 0.01$) significantly affected perceived externalization. If the reverberation was left unchanged (0% reduction), the externalization ratings decreased noticeably with increasing smoothing bandwidth above one ERB. In the case of most severe smoothing ($B = 64$), the median value of the externalization rating was about 1.4, corresponding to the sound image perceived as little externalized. When the reverberation was absent (100% reduction), the externalization ratings were low for all smoothing factors; only minor differences in externalization could be observed by reducing the spectral details in

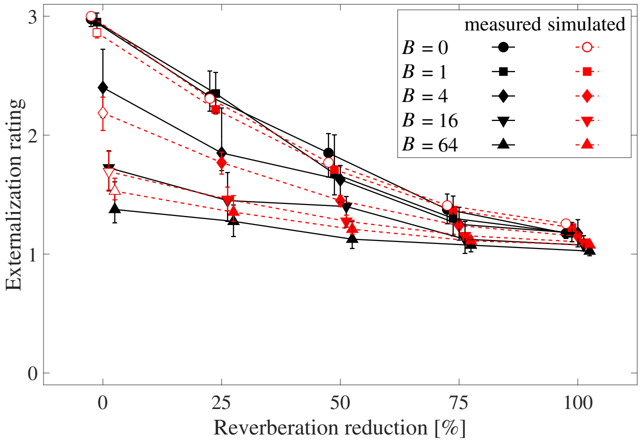

Figure 7. Median values of externalization ratings and the model predictions for different bilateral spectral smoothing ("*B*") and reverberation reduction levels across subjects. All other conventions are as in Figure 2.

the direct part. Note that the playback signals for this experimental condition (reverberation reduction factor of 100%) was the same as for the "bi" condition in Experiment D. Hence, the differences in externalization results were caused by the different experimental contexts with reference sounds either containing reverberation or not.

4 Model fitting and assessment

4.1 Fitting

The results of subjective measurements from a set of listening experiments confirmed that all three acoustic cues, i.e., spectral information, ILDs and reverberation did affect the degree of externalization. The sound could not be perceived as well-externalized when one of these cues was distorted. Part of the experimental results was used to calculate the weighted parameters in the externalization model (see Fig. 1), and the remaining experimental results were applied to evaluate the performance of the obtained externalized model.

If there is no deviation of acoustic cues between target and template signals, i.e., $\overline{\Delta\text{ILD}}$, $\overline{\Delta\xi}(w)$ and $\overline{\Delta\text{ILD}}_{\text{TSD}}$ are all zero, the externalization rating of this target signal should be 3. Therefore, the sum of coefficients a and c in the externalization mapping function (Eq. (9)) should be equal to 3. To further simplify the mapping function, the coefficient c was defined as the minimal externalization rating from our five experiments, i.e., c was equal to 1 (the sound was perceived as being at the ipsilateral ear). Therefore, the mapping parameters a and c were set to 2 and 1, respectively. The subjective results obtained in Experiments A (BB condition), B, C, and E ("$B = 0$" and "0% reverberation reduction" conditions) were used to adjust the weighting factors for these cues by fitting the model outcomes to the individual externalization ratings using the least-squares method (minimization of the summed square of residuals, MATLAB function *lsqnonlin*).

Table 2 shows the iterative steps applied to determine all the five model weightings ($b_{\text{ILD TSD}}$, b_{ILD}, b_{ξ}, w and b_{γ}) based on different experimental results. At each iterative step, the externalization results were used to determine the weighting factors individually for every subject, so that the confidence intervals for generic weighting factors could be estimated (see Table 3). Finally, the weighting factors were then averaged over the subjects and those generic values were applied during model assessment.

In the first step, the weighting factor $b_{\text{ILD TSD}}$ was determined based on the results from Experiment E for different reverberation reductions and unmodified spectral details ("$B = 0$", 5 data points per subject and parameter). Since the ILD TSD was the only acoustic cue that changed in this experimental condition (ILD and SGs were unchanged) and $b_{\text{ILD TSD}}$ can be decoupled from other four weighting factors (refer to Eq. (8)), the obtained factor $b_{\text{ILD TSD}}$ did not have to be re-optimized in the final step.

In Experiments A, B, and C, affected acoustic cue were coupled, at least weakly. Further, the reverberation-related weighting γ was unknown. Hence, b_{ILD}, b_{ξ} and w were first pre-optimized based on the externalization results in anechoic conditions (steps 2–4), where it can be assumed that $\overline{\Delta\text{ILD}}_{\text{TSD}}$ was small and γ was therefore close to one. These pre-optimal weightings obtained were then used as initial values to finally obtain the optimal weightings jointly (step 6).

In the second step, the b_{ILD} was calculated by fitting the $\overline{\Delta\text{ILD}}$ to the subjects' results of Experiment A under the condition "BB" (5 data points) without considering the reverberation-related weighting and SGs (γ and $\overline{\Delta\xi}(w)$ were set to one and zero, respectively).

In the third and fourth steps, the b_{ξ} and w were jointly optimized based on the externalization results in Experiments B and C: first, the weighting factors b_{ξ} were optimized for a predefined set of binaural weighting factors $w \in \{0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ based on the externalization results from Experiment B (5 data points) without considering the reverberation-related weighting ($\gamma = 1$); then, the best pair of w and b_{ξ} was selected as the argument for which the summed square of errors between the simulated and individually measured results across experimental conditions in Experiment C reached its minimum (10 data points).

In the fifth step, the weighting factor b_{γ} was adjusted based on the results in Experiment E under the "0% reverberation reduction" condition (5) data points) combined with pre-optimized weightings b_{ILD}, w and b_{ξ}.

Finally, the weighting factors, b_{ILD}, w, b_{ξ}, and b_{γ} were jointly re-optimized by minimizing the simulated and measured results (25 data points) from Experiments A ("BB" condition), B, C, and E ("0% reverberation reduction" conditions). The pre-optimized weighting factors were used as the initial values for the final optimization process.

The averaged weighting parameters across subjects and with their corresponding 95% CIs are calculated and listed in Table 3. Overall, the weighting factors obtained do not show large individual differences, especially the w and b_{γ}.

Table 2. The iterative steps of model fitting for each subject. N_d represents the number of data points per subject. Model variables and parameters are explained in Section 2. The results of obtained weighting factors are listed in Table 3.

Step	Experiment	Condition	N_d	Fitting parameters	Initial value	Fixed parameters
1	E	"$B = 0$"	5	$b_{\mathrm{ILD\ TSD}}$	Random	$b_{\mathrm{ILD}} = b_{\xi}(w) = b_{\gamma} = 0$
2	A	"BB"	5	b_{ILD}	Random	$b_{\mathrm{ILD\ TSD}}$ (step 1), $b_{\xi}(w) = 0,\ b_{\gamma} = 1$
3	B	–	5	b_{ξ} & w	b_{ξ} is random; $w \in \{0.5,0.6,0.7,0.8,0.9,1\}$	$b_{\mathrm{ILD\ TSD}}$ (step 1), b_{ILD} (step 2), $b_{\gamma} = 1$
4	C	"ipsi" &"contra"	10	b_{ξ} & w	Taken from step 3	$b_{\mathrm{ILD\ TSD}}$ (step 1), b_{ILD} (step 2), b_{γ} =1
5	E	"0% reverb. reduction"	5	b_{γ}	Random	$b_{\mathrm{ILD\ TSD}}$ (step 1), b_{ILD} (step 2), $b_{\xi}(w)$ (step 4)
6	A, B, C, E	"BB", "ipsi", "contra" & "0% reverb. reduction"	25	b_{ILD}, b_{ξ}, w and b_{γ}	Taken from steps 2, 3 and 4	$b_{\mathrm{ILD\ TSD}}$ (step 1)

4.2 Assessment

The averaged simulated results for each experiment are plotted with dashed lines in Figures 2–7. The mapped results are represented by open symbols. Experiments B ("LO" and "HI" conditions), D (all three conditions) and E ("$B = 1$", "$B = 4$", "$B = 16$", and "$B = 64$" conditions except for the 0% reverberation reduction) were used to test the performance of the model. To quantify how well the predicted results matched the measured externalization ratings, the normalized root mean square deviation (NRMSD) was calculated for each experimental condition and each subject [15]:

$$\mathrm{NRMSD} = \frac{1}{O_{\mathrm{range}}}\sqrt{\frac{\sum_{i=1}^{K}(P_i - O_i)^2}{K}}, \tag{16}$$

where P_i, and O_i denote predicted and measured externalization ratings for each condition, respectively. K is the number of experimental conditions. O_{range} is a normalization factor that was considered the maximal range of ratings in the experiment, i.e., O_{range} was equal to 3. The averaged NRMSD across subjects are presented in Table 4. The mapping and prediction errors of calculated results are shown in italic and bold, respectively. Overall, the averaged prediction errors were smaller than 10% (the overall accuracy was higher than 90%) for all experiments, corresponding to the high accuracy of the proposed externalization model.

5 Discussion

Five psychoacoustic experiments were conducted to show the relevance of three important acoustic cues in perceived externalization, and the experimental results were well predicted by our proposed model. Although no explicit control conditions were incorporated in the experimental design to rule out the inclusion of audible artifacts not related to externalization perception in the listeners' ratings, the following reasons strengthen our believe that ratings were at least primarily based on perceived externalization. First, listeners were clearly instructed on how to use the response scale and to ignore audible artifacts that do not affect their externalization perception. Second, part of experimental results are qualitatively consistent with previous studies, which did validate the effect of externalization loss by means of behavioral and neural (measured using electroencephalography) biases known to be elicited by approaching sounds [8]. Third, the model explaining the experimental results requires processing stages, such as imbalanced binaural weighting of spectral cues that are not clearly related to mere audibility.

Table 3. Mean weighting factors with ±95% CIs for different acoustic cues.

Factors	b_{ILD}	b_{ξ}	w	$b_{\mathrm{ILD\ TSD}}$	b_{γ}
Values	2.5 ± 1.1	1.7 ± 0.3	0.9 ± 0.1	2.8 ± 0.4	0.6 ± 0.1

In the following, each model component, interpretation of individual externalization results, and model limitations are discussed in details.

5.1 Model components

The subjective results of Experiment A revealed that the virtual sound source could not be perceived as well-externalized by maintaining only the correct spectral information of sound sources at each ear. Furthermore, the ILD expansion at low frequencies had more influences on perceived externalization than that at high frequencies. Thus, the deviation of the broadband ILD over the whole frequency range might not be appropriated to explain the externalization results by increasing the ILD at high frequencies. To demonstrate that, the deviation of broadband ILD was used to model the externalization result instead of the frequency-dependent relative ILD. As expected and despite re-optimization of b_{ILD}, high prediction errors were obtained for "LO" (NRMSD = 11.5%) and "HI" conditions (12.2%). Hence, the broadband ILD is

Table 4. Averaged prediction error (NRMSD) between the predicted and perceptual data. The mapping and prediction errors of calculated results are shown in italic and bold, respectively.

Experiment	Modification	Condition	Fitting parameter	NRMSD
A	ILD expansion	BB	b_{ILD}	*4.5%*
		LO		**4.8%**
		HI		**7.8%**
B	Spectral magnitude smoothing	ipsilateral, constant ILD	b_ξ & w	*4.9%*
C	ILD contrast compression	ipsi	b_ξ & w	*4.0%*
		contra	b_ξ & w	*4.0%*
D	Spectral magnitude smoothing	bi		**4.8%**
		ipsi		**3.7%**
		contra		**4.2%**
E	Reverberation reduction and spectral magnitude smoothing	High spectral detail ($B = 0$)	$b_{\mathrm{ILD\ TSD}}$ & b_γ	*3.3%*
		$B = 1$		**4.7%**
		$B = 4$		**6.4%**
		$B = 16$		**3.1%**
		Low spectral detail ($B = 64$)		**3.0%**

not appropriate for predicting the effect of frequency-specific ILD expansion on externalization.

In Experiment B, the magnitude spectra of HRTFs at both ears were changed, i.e., magnitude spectra of HRTFs at the ipsilateral ear were smoothed, and at the contralateral ear were adapted to keep the ILD constant within every frequency band. The results indicated that the spectral information of binaural sounds in each ear was relevant for perceived externalization. The degree of externalization was degraded by the deterioration of the spectral details in HRTFs, although the correct ILD information was maintained. This deterioration was generally consistent with previous studies [6] but more moderate in absolute terms presumably because the present experiment tested a more lateral sound source. Experiments C and D further investigated the relative influence of the spectral magnitude in HRTFs of the contralateral versus ipsilateral ear on perceived externalization. Hence, the contribution of the spectral information in HRTFs of the ipsilateral ear was more relevant for perceived externalization than that of the contralateral ear. The binaural weighting found here was more moderate than previously defined [24], such that the weight for the contralateral spectrum remained larger than zero even at the most extreme lateralization. In other words, the contralateral spectral details in HRTFs seem not to become negligible to reproduce well-externalized sound images.

The externalization results in Experiment D also showed that the degree of externalization was reduced by smoothing the spectral magnitude with bandwidths above one ERB. Smoothing the spectral magnitude of HRTFs caused not only a change of ILDs but also of spectral details in HRTFs. The calculated externalization ratings matched well to the measured externalization results. In order to test whether both cues were important to explain the results or whether a single cue would be sufficient, the externalization results were also simulated with using only one acoustic cue, i.e., only one cue weight was optimized and the other was forced to zero. For the SG-based model, the weighting factor for SG deviations, b_ξ, and the binaural weighting factor, w, were re-optimized based on the externalization results from Experiment C, since the influence of ILDs should not be taken into account for calculating these two factors. As a result, the re-calculated averages of b_ξ and w were 2.0 and 0.8, respectively. For the ILD-based model, the averaged b_{ILD} was equal to 2.4.

Figure 8 shows the predicted results for Experiment D based on either ILD deviations (dashed lines) or SG deviations (dash-dotted lines). For the SG-based model, the calculated externalization results for both "bi" and "ipsi" conditions matched well to the measured data with increasing bandwidths. However, the predicted results for the "contra" condition decreased only slightly across experimental conditions due to the low binaural weighting factor for contralateral SGs (cf. Experiment C). In the case of the ILD-based model, the trend of externalization results was comparable to the measured data with increasing bandwidths, but the relative externalization results between the "ipsi" and "contra" conditions were not consistent with the measured data. Additionally, the overall computed results were higher than the subjective data. These simulation results show that both monaural and interaural spectral cues are required to predict externalization results by smoothing the spectral magnitude of HRTFs. Additionally, the calculated weighting factors indicate that both cues contribute almost equally to externalization, since the values of b_{ILD} and b_ξ do not show large differences.

In Experiment E, if the reverberation was completely present (0% reverberation reduction condition), the degree of externalization depended on the smoothing levels (SG and ILD deviations) of the direct sound part, i.e., the externalization ratings decreased with increasing SG and ILD deviations. In contrast, if the reverberation was absent (100% reverberation reduction condition), the externalization ratings were low, even when the direct sound part was unmodified. This means that all three cues, SGs, ILDs, and ILD TSDs were relevant to perceived externalization, but only one of them alone was not sufficient to well

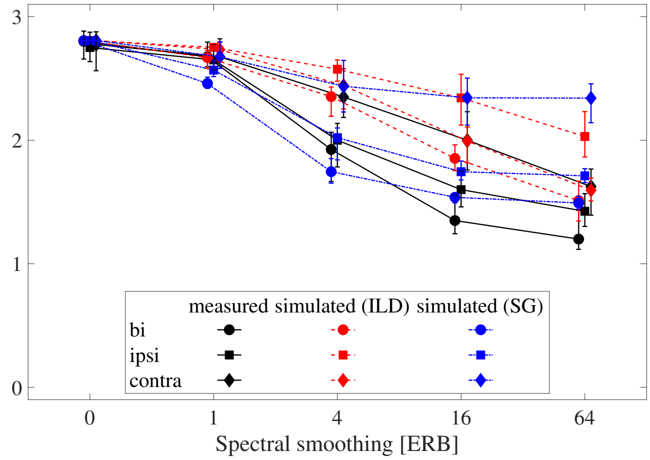

Figure 8. Median values of predicted externalization ratings using a single acoustic cue (ILD or SG deviations) by reducing the spectral details in HRTFs of both ears (condition “bi”), the ipsilateral ear (condition “ipsi”) and the contralateral ear (condition “contra”) across subjects. ILD- and SG-based prediction results are plotted with dashed and dash-dotted lines, respectively. All other conventions are as in Figure 2.

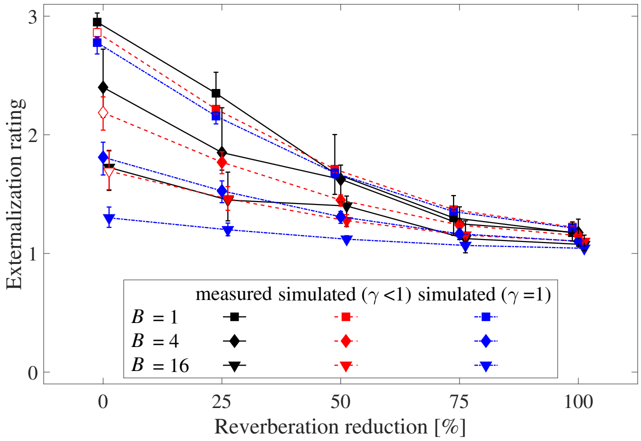

Figure 9. Median values of simulated externalization ratings with (dashed lines, “$\gamma < 1$”) and without (dash-dotted lines, “$\gamma = 1$”) the reduction term in the model for different spectral smoothing (“B”) and reverberation compression levels. All other conventions are as in Figure 2.

externalize sound images. Thus, the combination of these three cues was applied to build the externalization model, and the predicted results were well consistent with the measured externalization ratings.

In the present model, the influence of SG and ILD deviations was reduced by increasing the amount of reverberation-related cues. To point out how the influences of SGs and ILDs on modeled externalization ratings changed from an anechoic to a reverberant environment, the externalization for Experiment E was calculated without this reduction term ($\gamma = 1$), i.e., all other weighting factors were re-optimized without considering γ. Figure 9 shows the simulated externalization ratings for several experimental conditions (“$B = 1$”, “$B = 4$”, and “$B = 16$”) with and without the reduction term. Such model prediction clearly underestimated the perceptual data for bandwidths above one ERB, especially when the reverberation was completely present (0% compressed reverberation). The NRMSDs increased to 5.8%, 12.8%, 8.3% and 5.2% for “$B = 1$”, “$B = 4$”, “$B = 16$”, and “$B = 64$” conditions, respectively. This prediction degradation illustrates that the influences of spectral details and ILDs on perceived externalization are reduced when reverberation is present, and demonstrates that the proposed implementation is effective. Moreover, the values of the reduction term γ indicate that the contribution of spectral details to externalization in reverberant environments ($\gamma = 0.6$, refer to Eq. (7)) decreases by about 30% compared to anechoic environments (γ is about 0.9).

5.2 Individual differences

Due to listener-specific acoustics, applied modification procedures may differently affect the acoustic cues. For instance, smoothing the spectral magnitude of individual HRTFs caused different amounts of changes in ILDs and monaural spectral information. Consequently, the ILD could change more than the monaural spectral information for one subject and less for the other subject. As an example, the externalization ratings for “contra” and “ipsi” conditions in Experiment D are plotted in Figure 10 for exemplary subjects 2 and 3. Interestingly, they rated those two conditions qualitatively different. The median value of externalization results of subject 2 were higher for the “contra” condition than for the “ipsi” condition, whereas subject 3 rated both conditions as equivalent.

The objective measures, i.e., normalized ILD and SG deviations, were calculated for comparison with the subjective results and are shown in Figure 11. For subject 2 and large smoothing bandwidths, the deviation of normalized ILDs for the “contra” condition was slightly higher than for the “ipsi” condition, while the normalized SG deviation for the “ipsi” condition was much larger than for the “contra” condition. The high contribution of SG deviations thus dominated the opposed changes in ILD and resulted in higher externalization ratings for the “contra” condition. For subject 3, in contrast to subject 2, the deviations of normalized ILDs and SGs were more similar in size while still working in opposite directions, effectively equalizing the two conditions.

Smoothing the spectral magnitude of individual HRTFs led to different deviations in ILDs and spectral information and thus to different individual externalization ratings. Therefore, these two acoustic cues should be individually considered for predicting externalization. Overall, the individual prediction results (dashed lines in Fig. 10) are well consistent with the externalization results for both subjects. Hence, while idiosyncratic sensitivities may have modulated the spatial percept [35], the individual externalization results are highly correlated with the deviations of individual acoustic cues. This simulation result shows the importance of considering these marked

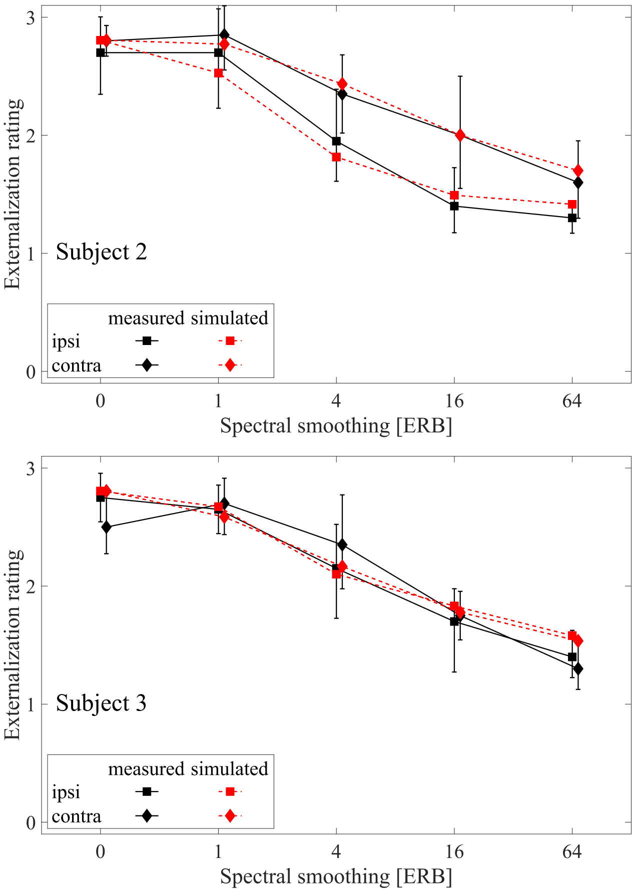

Figure 10. Median values of externalization ratings and the predicted results of subject 2 (upper panel) and subject 3 (lower panel) for smoothed spectral information in the HRTF ("contra" and "ipsi" conditions in Experiment D). All other conventions are as in Figure 2.

differences in individual acoustics for accurate predictions of externalization ratings.

5.3 Model limitations

A major limitation of the proposed model is that it does not take into account the influence of the sound source direction and different room acoustics. The weights for acoustic cues were calculated based on the analysis of a 90° sound source and two rooms (anechoic chamber and listening room) in the present study. However, if the incidence angle of the sound source would change, the absolute and relative influences of these cues on externalization ratings might be changed, as represented by the weights and the mapping parameters in the model. The directional dependence of the factor, w, for binaural weighting of monaural spectral cues can be assumed to be similar to the one established for sound localization in sagittal planes [24]. Baumgartner et al. [24] used a sigmoid function to determine the binaural weighting factors for an azimuth angle of φ: $w = (1 + \mathrm{e}^{(-\varphi/\Phi)})^{-1}$, where Φ is a scaling parameter, broadening the curve. In that study, the Φ was chosen as 13° to fit the outcomes of localization studies. For externalization

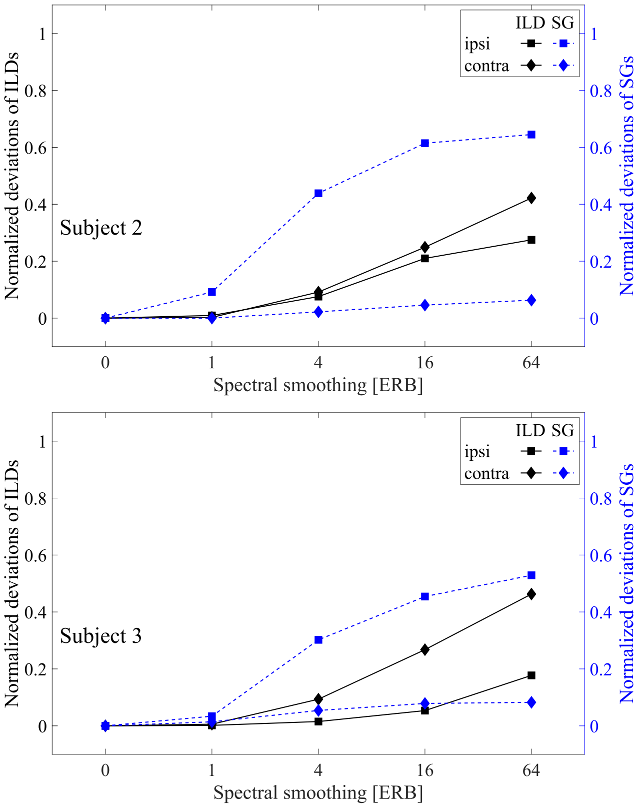

Figure 11. Deviation of normalized ILDs (solid lines) and SGs (dashed lines) of the individually synthesized binaural signals for subject 2 (upper panel) and subject 3 (lower panel) under different conditions.

perception, our present results suggest a more shallow dependence on the lateral angle ($\Phi = 41°$), in line with previous observations [18]. The offset, c, of the response mapping function is also expected to show directional dependency but this has not been formalized previously. As a first approximation one could assume a simple geometrical relationship based on a spherical head and internalized lateralization between the two ears (e.g., $c = |\sin(\varphi)|$). The validity of those interpolation methods needs to be tested in future experiments. Furthermore, the weighting factor, b_γ and $b_{\mathrm{ILD\ TSD}}$, depend on the defined reference ILD TSD in this study, and the existence of a general prior expectation of room acoustics is a matter of ongoing debate (c.f. [10, 36]).

The weighting factors in the model were calculated based on the subjective ratings, which are expected to be highly task-specific. In our experimental design, the loudspeaker was present at the measurement position to serve as a reference position, and the reference sounds from the loudspeaker were overtly presented to listeners. Calcagno et al. [37] showed that the visual information can potentially affect distance perception. Gil-Carvajal et al. [11] demonstrated that the auditory room-related cues between the recording and playback environment were more important than the visual room-related cues regarding distance perception. This issue may be critical for frontal sound sources, since lateral sources can be perceived as more exterznalized than frontal sources even without real source and visual cues as references [38].

Furthermore, our experiments and model investigations did not cover different sound source characteristics. Noise signal was chosen as the stimulus because it distributes the energy uniformly over frequencies and time, which is beneficial for the stable extraction of acoustic cues. Leclère et al. [38] reported that stimulus type had a small but significant influence on perceived externalization (speech was perceived as less externalized than noise and music), but it did not interact with BRIR individualization. Hence, the stimulus type seems not to affect the relevance of spectral details. However, it appears plausible to believe that specific stimuli may reduce the accessibility of certain acoustic cues for longer time periods and thus trigger an adaptation of perceptual weights. The proposed model may serve as a useful tool to analyze such potential adaption processes in future investigations.

Finally, the present study focused only on static listening; dynamic scenarios with head or source movements were not taken into account. Although the cues provided by dynamic scenarios not noticeably influence perceived externalization for a 90° sound source, they are highly relevant for more frontal sound sources [32, 39]. Hence, future model extensions should aim to incorporate the temporal integration processes necessary to evaluate such dynamic cues.

6 Conclusions

The present study investigated the role of monaural spectral cues, ILDs, and temporal fluctuations of ILDs in perceived externalization of a far-lateral sound source. The experimental results demonstrated that all three cues were perceptually relevant, and a single acoustic cue alone was not sufficient to well externalize the virtual sound image. Additionally, the contribution of the spectral information in HRTFs of the ipsilateral ear was more relevant for perceived externalization than that of the contralateral ear, but the contralateral spectral details in HRTFs should not be neglected to reproduce well-externalized sound images. Moreover, the influences of monaural spectral cues and ILDs on perceived externalization were reduced if reverberation is present.

A quantitative model was proposed to predict the degree of externalization based on the deviations between acoustic cues extracted from the target signals and expected templates. The predicted externalization ratings corresponded well to the obtained results. In contrast to the model proposed by Hassager et al. [17], not only frequency-dependent ILDs, but also monaural spectral cues and ILD temporal fluctuations were considered to predict perceived externalization. In the model, ILDs needed to be considered in a frequency-specific manner, and broadband ILDs were not sufficient for predicting perceived externalization. Further, the acoustic idiosyncrasies caused individual differences in the effects of spectral distortions on perceived externalization. The proposed externalization model can serve as a good starting point for further extensions, e.g., addressing various incidence angles, listening environments, and non-stationary scenes. Moreover, it may be used to generate hypotheses for experimental investigations in the future.

For the sake of open science, we incorporated the model implementation and model simulations into the Auditory Modeling Toolbox [40].

Acknowledgments

The authors would like to thank those who participated in the listening tests. R.B. was supported by the Austrian Science Fund (FWF): J3803-N30.

Conflict of interest

Authors declared no conflict of interests.

References

1. N.I. Durlach, A. Rigopulos, X.D. Pang, W.S. Woods, A. Kulkarni, H.S. Colburn, E.M. Wenzel: On the externalization of auditory images. Presence: Teleoperators & Virtual Environments 10, 2 (1992) 251–257.
2. J. Blauert: The Technology of Binaural Listening. Springer, Heidelberg, Berlin, 2013.
3. F.L. Wightman, D.J. Kistler: Headphone simulation of free–field listening. I: Stimulus synthesis. Journal of the Acoustical Society of America 85 (1988) 858–867.
4. M. Vorländer: Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality. Springer, Heidelberg, Berlin, 2008.
5. A.J. Kolarik, B.C.J. Moore, P. Zahorik, S. Cirstea, S. Pardhan: Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss. Attention, Perception, & Psychophysics 78, 2 (2016) 373–395.
6. W.M. Hartmann, A. Wittenberg: On the externalization of sound images. Journal of the Acoustical Society of America 99 (1996) 3678–3688.
7. A. Kulkarni, H.S. Colburn: Role of spectral detail in sound-source localization. Nature 3960, 6713 (1998) 747.
8. R. Baumgartner, D.K. Reed, B. Tóth, V. Best, P. Majdak, H. S. Colburn, B. Shinn-Cunningham: Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias. Proceedings of the National Academy of Sciences 1140, 36 (2017) 9743–9748.
9. D.R. Begault, E.M. Wenzel, M.R. Anderson: Direct comparison of the impact of head tracking, reverberation, and individualized Head-related transfer functions on the spatial perception of a virtual speech source. Journal of the Audio Engineering Society 49 (2001) 904–916.
10. S. Werner, F. Klein, T. Mayenfels, K. Brandenburg: A summary on acoustic room divergence and its effect on externalization of auditory events, in 8th International Conference on Quality of Multimedia Experience (QoMEX), IEEE. 2016, pp. 1–6.
11. J.C. Gil-Carvajal, J. Cubick, S. Santurette, T. Dau: Spatial hearing with incongruent visual or auditory room cues. Scientific Reports 6 (2016) 37342.

12. F. Völk: Externalization in data-based binaural synthesis: Effects of impulse response length, in Tagungsband Fortschritte der Akustik-DAGA 2009. 2009, pp. 1075–1078.
13. R. Crawford-Emery, H. Lee: The subjective effect of BRIR length perceived headphone sound externalisation and tonal colouration, in 136th Audio Engineering Society Convention. 2014.
14. J. Catic, S. Santurette, T. Dau: The role of reverberation-related binaural cues in the externalization of speech. Journal of the Acoustical Society of America 138 (2015) 1154–1167.
15. S. Li, R. Schlieper, J. Peissig: 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. Journal of the Acoustical Society of America 1440, 2 (2018) 966–980.
16. B.G. Shinn-Cunningham, N. Kopco, T.J. Martin: Localizing nearby sound sources in a classroom: Binaural room impulse responses. Journal of the Acoustical Society of America 117 (2005) 3100–3115.
17. H.G. Hassager, F. Gran, T. Dau: The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment. Journal of the Acoustical Society of America 139 (2016) 2992–3000.
18. S. Li, R. Schlieper, J. Peissig: The role of reverberation and magnitude spectra of direct parts in contralateral and ipsilateral ear signals on perceived externalization. Applied Sciences 90, 3 (2019) 460.
19. R. Baumgartner, P. Majdak, B. Laback: Modeling the effects of sensorineural hearing loss on sound localization in the median plane. Trends in Hearing 20 (2016) 1–11.
20. G. Plenge: Über das problem der im-kopf-lokalisation (The problem of in-head lokalization). Acustica 26 (1972) 241–252.
21. C. Mendonça: A review on auditory space adaptations to altered head-related cues. Frontiers in Neuroscience 8 (2014) 219.
22. R.F. Lyon: All-pole models of auditory filtering, in Diversity in Auditory Mechanics Lewis ER, Long GR, Lyon RF, Narins PM, Steele CR, Hecht-Poinar E, Editors, World Scientific Publishing, Singapore. 1997, pp. 205–211.
23. B.R. Glasberg, B.C.J. Moore: Derivation of auditory filter shapes from notched-noise data. Hearing Research 47 (1990) 103–138.
24. R. Baumgartner, P. Majdak, B. Laback: Modeling sound-source localization in sagittal planes for human listeners. Journal of the Acoustical Society of America 2, 1360 (2014) 791–802.
25. H. Wallach, E.B. Newman, M.R. Rosenzweig: The precedence effect in sound localization. The American Journal of Psychology 62 (1949) 315–336.
26. J. Braasch: A precedence effect model to simulate localization dominance using an adaptive, stimulus parameter-based inhibition process. Journal of the Acoustical Society of America 1, 1340 (2013) 420–435.
27. R. Baumgartner, P. Majdak: Decision making in auditory externalization perception. Preprint (2020). bioRxiv 2020.04.30.068817, https://doi.org/10.1101/2020.04.30.068817.
28. S. Müller, P. Massarani: Transfer-function measurement with sweeps. Journal of the Audio Engineering Society 49 (2001) 443–471.
29. S. Li, J. Peissig: Measurement of head-related transfer functions: A review. Applied Sciences 10, 14 (2020) 5014.
30. A. Kulkarni, S.K. Isabelle, H.S. Colburn: On the minimum-phase approximation of head-related transfer functions, in Proceedings of 1995 Workshop on Applications of Signal Processing to Audio and Accoustics, IEEE. 1995, pp. 84–87.
31. Z. Schärer, A. Lindau: Evaluation of equalization methods for binaural signals, in 126th Audio Engineering Society Convention. 2009.
32. W.O. Brimijoin, A.W. Boyd, M.A. Akeroyd: The contribution of head movement to the externalization and internalization of sounds. PLoS One 8 (2013) 1–12.
33. F.L. Wightman, D.J. Kistler: Monaural sound localization revisited. Journal of the Acoustical Society of America 1010, 2 (1997) 1050–1063.
34. A. Kohlrausch, J. Breebaart: Perceptual (ir) relevance of HRTF magnitude and phase spectra, in 110th Audio Engineering Society Convention. 2001.
35. P. Majdak, R. Baumgartner, B. Laback: Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization. Frontiers in Psychology 5 (2014) 319.
36. F. Klein, S. Werner, T. Mayenfels: Influences of training on externalization of binaural synthesis in situations of room divergence. Journal of the Audio Engineering Society 65, 3 (2017) 178–187.
37. E.R. Calcagno, E.L. Abregu, M.C. Egua, R. Vergara: The role of vision in auditory distance perception. Perception 41, 2 (2012) 175–192.
38. T. Leclère, M. Lavandier, F. Perrin: On the externalization of sound sources with headphones without reference to a real source. Journal of the Acoustical Society of America 146, 4 (2019) 2309–2320.
39. E. Hendrickx, P. Stitt, J.C. Messonnier, J.M. Lyzwa, B.F.G. Katz, C. Boishéraud: Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis. Journal of the Acoustical Society of America 141 (2017) 2011–2023.
40. P. Søndergaard, P. Majdak: The auditory modeling toolbox, in The Technology of Binaural Listening Blauert J, Editors, Springer, Berlin, Heidelberg. 2013, pp. 33–56.

Cite this article as: Li S, Baumgartner R & Peissig J. 2020. Modeling perceived externalization of a static, lateral sound image. Acta Acustica, **4**, 21.