


# Sound Externalization: A Review of Recent Research

Trends in Hearing  
Volume 24: 1–14  
© The Author(s) 2020  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2331216520948390  
journals.sagepub.com/home/tia  


Virginia Best<sup>1</sup> , Robert Baumgartner<sup>2</sup>, Mathieu Lavandier<sup>3</sup>,  
Piotr Majdak<sup>2</sup>  and Norbert Kopčo<sup>4</sup>

## Abstract

Sound externalization, or the perception that a sound source is outside of the head, is an intriguing phenomenon that has long interested psychoacousticians. While previous reviews are available, the past few decades have produced a substantial amount of new data. In this review, we aim to synthesize those data and to summarize advances in our understanding of the phenomenon. We also discuss issues related to the definition and measurement of sound externalization and describe quantitative approaches that have been taken to predict the outcomes of externalization experiments. Last, sound externalization is of practical importance for many kinds of hearing technologies. Here, we touch on two examples, discussing the role of sound externalization in augmented/virtual reality systems and bringing attention to the somewhat overlooked issue of sound externalization in wearers of hearing aids.

## Keywords

sound image externalization, in-head localization, spatial perception, virtual acoustics, distal attribution

Received 8 May 2020; revised 11 June 2020; accepted 18 July 2020

This review is concerned with the perceptual externalization of auditory events, that is, the perception that a sound source is outside of the head. While the phenomenon of perceptual externalization already intrigued the pioneers of psychoacoustics (Weber, 1849), the majority of research on this topic came after the introduction of the binaural mode of sound presentation via headphones in the late 20th century. Most sounds originate from a physical source located outside of the head and, when received with open ears in the natural world, are generally perceived to be externalized. When sounds are presented diotically via headphones, however, the perception is generally that the sound images are inside the head, or internalized (Jeffress & Taylor, 1961). Occasionally, there can be a breakdown in externalization even for sounds presented via loudspeakers (Brimijoin et al., 2013; Plenge, 1974; Toole, 1970).

Durlach et al. (1992) reviewed the cues for externalization, motivated by the then relatively new technology of virtual auditory displays, for which the externalization aspect is crucial but easily disrupted. Blauert (1997, Chapter 2) also reviewed the evidence for several

explanations of what he called “inside-the-head locatedness.” Since that time there has been a wealth of new empirical data collected, and our understanding of the factors driving sound externalization has advanced considerably. Moreover, the current pervasiveness of headphones and the emergence of many new kinds of ear-worn technologies (Carlile et al., 2017; Zeng, 2016) suggest that the broad relevance of this topic continues to grow.

<sup>1</sup>Department of Speech, Language and Hearing Sciences, Boston University, Boston, MA, USA

<sup>2</sup>Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria

<sup>3</sup>Univ Lyon, ENTPE, Laboratoire Génie Civil et Bâtiment, Vaulx-en-Velin, France

<sup>4</sup>Institute of Computer Science, Faculty of Science, Pavol Jozef Šafárik University, Košice, Slovakia

### Corresponding Author:

Virginia Best, Department of Speech, Language and Hearing Sciences, Boston University, 635 Commonwealth Avenue, Boston, MA 02215, USA.  
Email: ginbest@bu.edu



Motivated by these developments, in this review, we focus on recent advances in both basic and applied research on sound externalization. The review includes all of the psychoacoustical studies returned by searching the PubMed and Web of Science databases with keywords “sound” and “externalization,” as well as other relevant references identified via snowball sampling and personal knowledge where appropriate (Greenhalgh & Peacock, 2005). In this review, we first discuss the definitions of sound externalization and describe the different ways in which it is measured. Then, we summarize the physical and contextual cues contributing to the percept of externalized sounds. Finally, we discuss augmented and virtual reality (A/VR) applications that deal with realistic sound reproduction and review data related to the relatively overlooked issue of sound externalization in wearers of hearing aids.

## Defining and Measuring Externalization

### *The Challenge of Defining Externalization*

The main prerequisite for externalization is generally thought to be an “ear-adequate” signal (Plenge, 1974), meaning that a binaural signal will be externalized as long as it provides all of the cues that would be available to the ears in some natural listening situation, encompassing properties of the source, the listener, and the environment. However, a realistic binaural sound reproduction does not need to be indistinguishable from the natural one. To be perceived as realistic, the physical cues simply need to create a single impression that fits within a listener’s experiences and beliefs about the world. Thus, it is not surprising that externalization has been related to attributes such as convincingness, presence, and realism (e.g., Durlach et al., 1992; Hartmann & Wittenberg, 1996; Simon et al., 2016). Sound externalization has also been discussed in a related literature on “distal attribution,” which is concerned with how sensory experiences are referred to external objects rather than to the self (e.g., Loomis, 1992). The idea that sound *externalization* relies on a correspondence between the signals received by a listener and that listener’s expectations about the spatial attributes of the acoustic environment has a complementary form, which says that *internalization* is the result of a violation of expectations. These two ideas will be a recurring theme in the sections that follow and provide a unifying framework for understanding a number of seemingly disparate findings.

Another interesting debate around the definition of sound externalization concerns its relationship to auditory distance perception. Most commonly, sound externalization is assumed to share the same continuum as distance perception, whereby the center of the head

represents the minimum possible distance of an auditory image. This overlap is strikingly clear in the interchangeable methods that have been used to measure each percept (see next section). Durlach et al. (1992, p. 251) state that externalization

...is a matter of degree: a source can appear far outside the head, near the border between inside and outside the head, or well inside the head. Thus, externalization can be thought of as a crude representation of subjective distance.

Similarly, Hartmann and Wittenberg (1996) interpreted the fact that they were able to systematically move a source from inside the head to outside as evidence for a continuous dimension that encompassed locations on both sides of the skull. From a physiological perspective, externalized sounds yield increased activity in the cortical area planum temporale (Callan et al., 2013; Hunter et al., 2003) that is known to be involved in distance perception (Kopčo et al., 2012).

Internalization may also represent a failure of distance perception. For example, it has been shown that a sound source located at a specific azimuth and near-field distance gives rise to a specific combination of binaural cues (Shinn-Cunningham et al., 2000). Thus, it follows that any stimulus presented with realistic binaural cues should produce an externalized percept with a specific azimuth and distance. One possibility is that sounds with unnatural cue combinations, because they cannot be mapped to any real external source location, tend to be internalized. While compelling, this idea has not been explored in detail.

Despite the general intermixing of perceived distance and degree of externalization, it remains a possibility that they represent distinct dimensions. Certainly, externalization has a clear dependence on binaural cues (as will be discussed later), while distance perception is dominated by monaural cues (for reviews, see Kolarik et al., 2016; Zahorik et al., 2005). Moreover, distance perception has been commonly studied under conditions in which externalization is known to be weak (e.g., a frontal source or diotic listening), suggesting that distance judgments are possible for poorly externalized (or even internalized) sources (e.g., Bidart & Lavandier, 2016; Kopčo et al., 2020). There are also reports that poor externalization tends to be accompanied by perceptual changes beyond reductions in distance, such as diffuseness or split images (e.g., Catic et al., 2013; Cubick et al., 2018; Hartmann & Wittenberg, 1996; Hassager, Wiinberg, et al., 2017). Hence, it remains unclear whether externalization and distance perception are distinct or whether they form part of the same continuum. What is clear is that it is difficult to disentangle them based on most of the common experimental approaches.

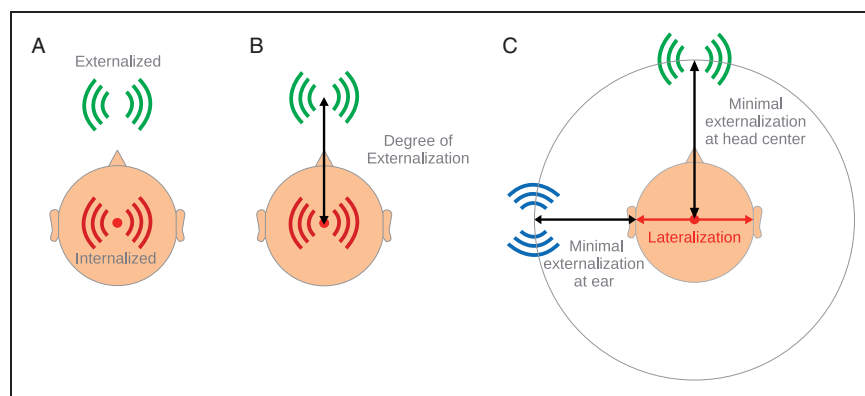
## Approaches to Measuring Externalization

While a handful of studies have examined sound externalization using stimuli presented over loudspeakers, the majority of studies have used stimuli presented over headphones, or a combination of these two. Headphone studies commonly make use of virtual auditory space techniques based either on head-related transfer functions (HRTFs), describing the acoustic filtering of a listener's pinnae, head, and body, or binaural room impulse responses (BRIRs), which additionally include the acoustics of the listener's environment.

Sound externalization has been quantified using a variety of psychophysical tasks. A simple method is to ask listeners to give a binary judgment indicating whether a sound image is perceived to be inside the head or outside the head (e.g., Brimijoin et al., 2013; see Figure 1A). More indirectly, listeners can be asked to discriminate between sounds presented under two different conditions, one in which the sounds should be fully externalized and one in which externalization may be disrupted to various degrees (Hartmann & Wittenberg, 1996; Kulkarni & Colburn, 1998). The most common method is to provide a continuous externalization scale (see Figure 1B), in which one end point of the scale is the center of the head and the other end point is some fixed distance in external space, often marked via a silent reference loudspeaker (e.g., Catic et al., 2013; Gil-Carvajal et al., 2016; Hartmann & Wittenberg, 1996; Hassager et al., 2016; Kim & Choi, 2005). However, a wide variety of labels have been used with continuous scales, especially for intermediate locations on the scale. For example, Catic et al. (2013, 2015) used terms only related to distance ("closer to me," "closer to the loudspeaker"), whereas Hartmann and Wittenberg (1996) also included compactness ("very diffuse") and mislocalization ("at the wrong place") in their 4-point scale. Continuous scales in inches have also been used (Begault & Wenzel, 1993; Begault et al., 2001). Others did not refer to distance and did not provide intermediate

labels but asked the participants to give a continuous rating ranging from "perceived inside the head" to "completely externalized" (Boyd et al., 2012; Leclère et al., 2019). Relative scales derived from ranking orders have been used as an alternative approach to direct scaling (e.g., Yuan et al., 2017). As a more multi-dimensional approach to evaluate spatial perception, Toole (1970) asked listeners for ratings of the direction and distance of externalized sources and for the percentage of the total sound that is located inside the head for internalized images. A related approach is to let listeners sketch sound source positions using a template that includes the head and some reference points in external space (e.g., Cubick et al., 2018; Hassager, May, et al., 2017; Hassager, Wiinberg, et al., 2017; Robinson & Xiang, 2013), although this approach has not been extended to include a quantification of the degree of externalization based on the sketches.

It is worth noting that some methods may introduce errors or biases that can affect the conclusions that are drawn and complicate comparisons across conditions and studies. Take for example a study in which sounds are presented from the front and from the side, under conditions in which externalization is expected to vary. In this case, the frontal sounds will likely collapse to the center of the head, whereas the lateral sounds will likely collapse to the ipsilateral ear (see Figure 1C). If the same continuous externalization scale is used for both cases, with the center of the head as an end point, a listener may not be inclined to give a rating of zero for the lateral sounds, which may introduce a bias toward higher externalization ratings. Another issue concerns the use of a visual reference, such as a loudspeaker, as is commonly used. If listeners are asked to judge externalization using such a reference, their judgments may be influenced by characteristics of the sound image such as its precise location in extrapersonal space. For example, a binaural signal that preserves externalization but distorts perceived distance may never be given a "perfect"



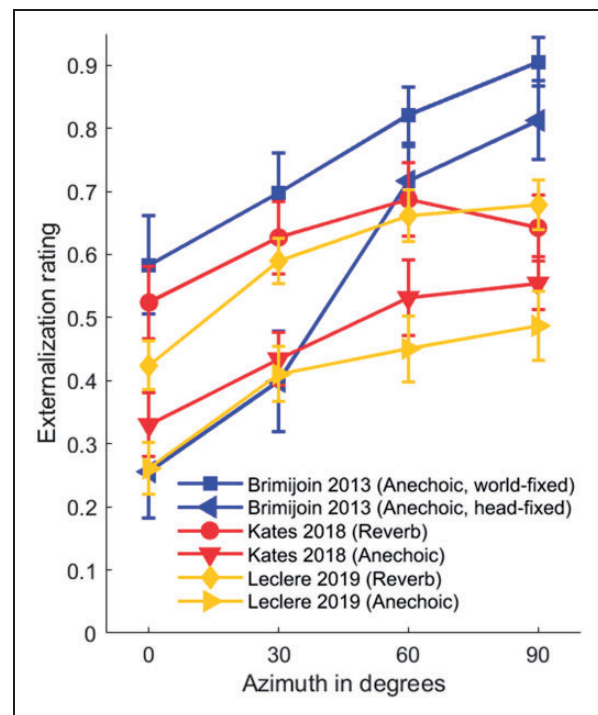
**Figure 1.** Sound externalization (A) as a binary decision, (B) as a continuum, and (C) depending on lateral position.

externalization rating even if is clearly perceived outside of the head (Leclère et al., 2019). The influence of visual information is discussed further below. Another issue with some approaches is that response averaging (by the listener or the experimenter) can result in the misrepresentation of well-externalized but spatially ambiguous sound images (Durlach et al., 1992; Werner et al., 2016). For example, in the sketching paradigm, if participants experience front-back confusions and their sketches depict a location in front on some trials but a location in the back on other trials, averaging may give the impression of a larger image with a more internalized center.

## Cues for Externalization

### Acoustic Cues in the Direct Sound

One line of research on this topic has concentrated on the perception of sounds presented in anechoic space, or under experimental conditions in which the influence of reverberation is largely factored out. While somewhat artificial, the anechoic case has proven to be quite informative because externalization breaks down easily and is sensitive to various stimulus manipulations. A basic observation that has been made under these conditions is that externalization is most robust when there is a difference between the ears. Specifically, sound sources located close to the median plane, where interaural differences are small, are more likely to be internalized than those located off to the side, where interaural differences are larger (e.g., Begault & Wenzel, 1993; Brimijoin et al., 2013; Kates et al., 2018; Kim & Choi, 2005; Leclère et al., 2019). In-the-head localization is also more prevalent when noise is presented from multiple loudspeakers, thus creating a more diffuse sound field with reduced interaural differences (Toole, 1970). Figure 2 shows data from three studies that obtained externalization ratings as a function of sound source azimuth (Brimijoin et al., 2013; Kates et al., 2018; Leclère et al., 2019). There is a clear tendency for externalization ratings to increase with sound source laterality. The tendency for greater externalization with more lateral sources may arise in part because of the response bias mentioned earlier (see Figure 1C), although there is no direct evidence for this. Another way to understand these effects is in terms of near-field distance perception (Shinn-Cunningham et al., 2000). Specifically, while unique combinations of interaural time difference (ITD) and interaural level difference (ILD) correspond to off-midline sources at different distances (i.e., external locations), no such unique mapping exists when the ITD and ILD are close to zero. It is possible that this ambiguity causes distance perception to break down (and lead to



**Figure 2.** Mean externalization ratings from three studies showing the tendency for increased externalization with sound source laterality. Brimijoin et al. (2013,  $N = 11$ ): Ratings correspond to the proportion of externalized responses obtained using a binary judgment and are averaged across fullband and lowpass conditions. Small head movements were allowed, and the sources were either fixed in azimuth (“world-fixed”) or moved along with the head movements (“head-fixed”). Kates et al. (2018,  $N = 20$ ): Ratings were obtained using a 100-point scale with visual references and are shown separately for anechoic and reverberant conditions. Leclère et al. (2019,  $N = 21$ ): Ratings were obtained using a continuous percentage scale with eyes closed and are shown separately for anechoic and reverberant conditions (the latter averaged across four rooms). Error bars show standard errors of the mean.

internalization) more frequently for sources near the midline.

Hartmann and Wittenberg (1996) used a headphone synthesis technique to examine the role of binaural cues in sound externalization. By imposing natural combinations of ITD and ILD on harmonic complexes presented over headphones, they were able to create virtual sounds that were indistinguishable from loudspeaker-produced sounds. Then, they manipulated the binaural cues and examined the effect on both discrimination and externalization, after controlling for any changes in horizontal position. They reported an increase in listeners’ ability to discriminate real from virtual sources when ITDs and ILDs deviated from their natural values, and externalization ratings showed a large proportion of in-the-head responses in the extreme cases of ITDs and ILDs corresponding to opposite sides of the head. They further

concluded that ITDs contributed to sound externalization for frequencies below 1 kHz, while ILDs contributed at all tested frequencies.

Hartmann and Wittenberg also suggested that a natural spectral shape at the ear was important to achieve full externalization, implicating pinna cues in addition to plausible binaural cues. This idea was examined more closely by Boyd et al. (2012), who obtained externalization ratings for speech presented in a sound-treated classroom. They showed that externalization ratings were reduced by mixing HRTF-filtered stimuli with equivalent stimuli measured using a pair of microphones with the head absent (i.e., containing ITDs but no head or pinna-related cues). By spectrally smoothing the HRTFs, Kulkarni and Colburn (1998) suggested that fine spectral details are not critical for broadband noises to be convincingly externalized. On the other hand, Baumgartner et al. (2017) spectrally flattened the HRTFs at the two ears systematically and showed that sound images gradually collapsed toward the head. Hassager et al. (2016) also showed that spectral smoothing of the direct parts of BRIRs, which roughly correspond to free-field HRTFs, degraded the externalization of band-limited noises (< 6 kHz). Li et al. (2019) replicated the basic results of Hassager et al. and demonstrated that spectral detail in the direct sound at the ipsilateral ear became increasingly important (relative to the contralateral ear) for more laterally located sound sources. An illustration of the effects of spectral smoothing and spectral flattening on HRTFs is provided in Figure 3.

The effect of modified spectral cues can also be related to the problem of individualization of HRTFs. Nonindividualized HRTFs, such as those measured on a dummy head or on another person, usually do not resemble a listener's own HRTFs, and in that sense stimuli derived from them are not natural or ear-adequate. A handful of studies have examined whether the individualization of HRTFs is critical for an externalized percept, with somewhat inconsistent conclusions. Kim and Choi (2005) reported that a few of their listeners benefited from HRTF individualization for the externalization of broadband noises. Consistent improvements in externalization were reported by Werner et al. (2016) when BRIRs were individualized versus generic, especially for the poorly externalized midline locations (which were also prone to front-back confusions). While they did not include individualized HRTFs, Simon et al. (2016) established that listeners perceive differences in externalization between different nonindividualized HRTFs. On the other hand, Begault et al. (2001) found no significant difference between externalization ratings of speech stimuli for individualized or generic HRTFs. Using noise stimuli presented in a number of reverberant rooms, Leclère et al. (2019) also found no effect of

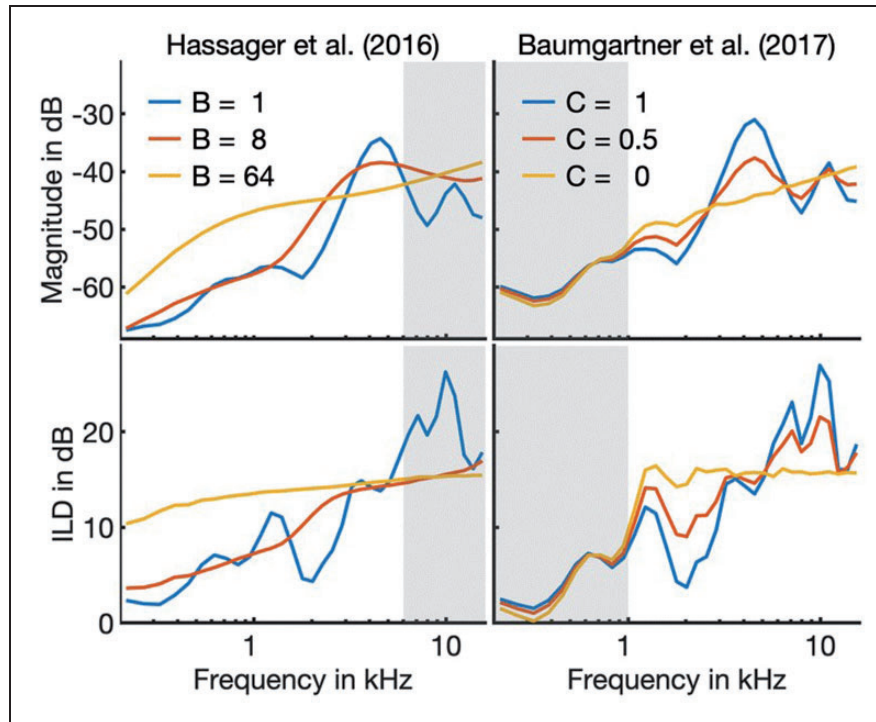
individualization of BRIRs on externalization. Taken together, the evidence suggests that listener-specific spectral cues may not be critical for externalization in more realistic listening situations.

### *Reverberation-Related Cues*

Reverberant sounds are more likely than anechoic sounds to be perceived as externalized (e.g., Begault et al., 2001; Leclère et al., 2019; Toole, 1970). Figure 2 shows externalization ratings obtained under anechoic and reverberant conditions in two different studies (Kates et al., 2018; Leclère et al., 2019). Interestingly, a comparison of these two data sets reveals a remarkable correspondence between ratings obtained under similar conditions but using different scales and different subject groups. Other studies using truncated BRIRs suggest that externalization is improved with the inclusion of up to 80 ms of the reverberant tails, with no further improvement for longer tails (e.g., Begault et al., 2001; Catic et al., 2015). It has also been suggested that high-frequency sounds are less well externalized than low-frequency sounds (e.g., Levy & Butler, 1978) because high-frequency energy is better absorbed by a room, resulting in less reverberation.

Studies investigating the role of pinna cues concluded that spectral detail in the reverberant sound is less important for externalization than spectral detail in the direct sound (Hassager et al., 2016; Jiang et al., 2020). On the other hand, many studies have shown that binaural information in the reverberant signal is critical for externalization. Catic et al. (2015) found that presenting the reverberant part of their speech signals diotically caused a clear reduction in externalization ratings. Similarly, Leclère et al. (2019) suggested that reverberation only improves externalization when it creates interaural differences; adding diotic reverberation had very little effect on externalization ratings. The results of Catic et al. (2015) further suggest that binaural cues from reflections are more important for externalization when the direct sound itself contains only weak binaural cues (e.g., frontal sources) and less important when the direct sound contains larger interaural differences (e.g., lateral sources).

Another focus of research has been to understand specifically which binaural cues are responsible for the externalization of reverberant sounds. While early studies highlighted the importance of ITDs (e.g., Levy & Butler, 1978), the focus in more recent studies has been on ILDs and interaural coherence (IC). For example, Catic et al. (2013) used BRIRs to understand the binaural cues contributing to the externalization of speech presented in a room. Their data suggested that fluctuations over time in the interaural cues may be critical. For sounds with energy higher than 1 kHz, they showed



**Figure 3.** Effects of spectral smoothing methods on an example HRTF at 60° azimuth. Top row: monaural spectrum at the ipsilateral ear. Bottom row: difference between the left- and right-ear spectra (or the ILD). Spectral representations were obtained by filtering the impulse response with a gammatone filter bank (Lyon, 1997) and calculating the root mean square amplitude within each band. Left column: smoothing according to Hassager et al. (2016) where  $B$  is the spectral smoothing bandwidth relative to one  $ERB_N$  (Glasberg & Moore, 1990). Right column: smoothing according to Baumgartner et al. (2017) where  $C$  is a scaling factor that is applied to spectral magnitudes. Gray-shaded areas indicate frequency ranges not tested in the respective experiments. The HRTF is from subject S01 in Baumgartner et al. (2017). ILD = interaural level difference.

qualitative correspondence between the standard deviation of ILDs across time and externalization ratings. In a second study, using a broader range of conditions and stimuli, Catic et al. (2015) concluded that temporal variations in both ILDs and IC could drive externalization ratings. Leclère et al. (2019) also found a good correspondence between externalization ratings and IC fluctuations (but also overall IC) in their stimuli. Li et al. (2018, 2019) investigated these cues further by selectively modifying the BRIR of the ipsilateral or contralateral ear (by truncating or attenuating the reverberant energy). They showed that, for a lateral sound source, externalization is particularly dependent on the characteristics of the reverberation reaching the contralateral ear, where fluctuations are in general larger.

Although the effect of reverberation on externalization is robust, it is still the case that the received signals must be plausible, natural, and consistent with the listener's expectations for full externalization. Early studies suggested that externalization can break down if the received signals do not match the listener's "stored information" about a room (Plenge, 1974), and in audio engineering, a similar phenomenon has been

referred to as the "room divergence effect" (e.g., Klein et al., 2017; Werner et al., 2016).

### *Multimodal Factors: Head Movements and Vision*

Dynamic listening situations are far more common than static situations in real-world listening due to the movement of sound sources and/or the listener. Behavioral experiments support the idea that head movements are essential for accurate sound localization and especially for resolving front from back (Brimijoin & Akeroyd, 2012; Wallach, 1940; Wightman & Kistler, 1999). Similarly, a sound source that moves relative to the head during head rotation is likely to be perceived as externalized, even if pinna-related information is absent (Loomis et al., 1990). Conversely, a source that "follows" the head is likely to collapse and be internalized, independent of whether headphones or loudspeakers are used for presentation (Brimijoin et al., 2013). Figure 2 provides a direct comparison between a listening condition with sound sources following the head movements ("head-fixed") and sound sources that stay in place when the head moves ("world-fixed").

Interestingly, it has also been reported that the improvement in externalization afforded by dynamic cues can persist in time beyond the dynamic exposure (Hendrickx et al., 2017b). The fact that head movements provide a larger benefit for centrally located sounds (Brimijoin et al., 2013; Hendrickx et al., 2017b) suggests that the benefit may come from access to binaural differences that support externalization. The finding that self-initiated movements enhance externalization more than source movements (Hendrickx et al., 2017a) further suggests that active control and the related sensory predictability are important.

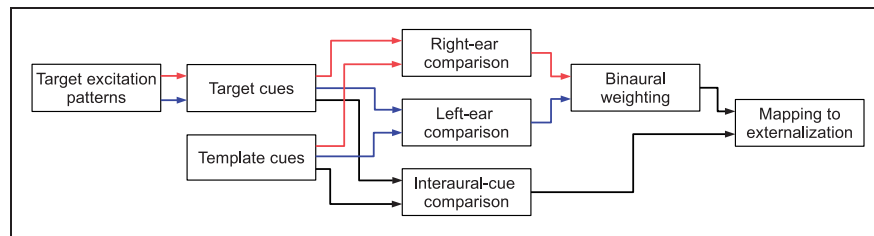
Vision is another modality that is well known to influence auditory spatial perception (e.g., Jones & Kabanoff, 1975; Majdak et al., 2010) and may also affect externalization. It has been speculated that a contributing factor to the breakdown of externalization during headphone presentation might be the lack of congruent visual information supporting the existence of an externalized sound source. To test this idea, Werner et al. (2016) used BRIRs to present simulated target sounds at various locations around the listener in two different rooms. They reported an average increase in externalized ratings in a group of listeners who could see the simulated room and loudspeakers compared with a group of listeners who were tested in darkness. Udesen et al. (2015) and Werner et al. (2016) examined the effect of incongruent visual information about the listening environment. They simulated target sounds in one room and measured externalization of these sounds while participants were situated in the same room, or in a different room. They found reductions in externalization ratings in the incongruent room. Similarly, Gil-Carvajal et al. (2016) found reduced externalization ratings for speech when the playback room did not match the simulated room. They divided their participants into two groups: One group had visual awareness, and the other group had auditory awareness of the testing room. Specifically, the “visual” group could see the room but were given no sounds besides the test stimuli. The “auditory” group were blindfolded but could hear a loudspeaker emitting bursts of noise. By comparing these groups, they concluded that externalization responses were primarily affected by incongruent auditory cues obtained in the playback room, not the visual impression of the room. Together, these results are compatible with the idea that externalization is most robust when all of the available cues are consistent with the perceiver’s expectations; incongruencies within or across modalities increase the chances of a breakdown in externalization.

### *Models of Sound Externalization*

Our understanding of sound externalization is closely linked with the concepts of sound localization. Plenge (1972) introduced a conceptual model of sound

localization that consists of a long-term memory and a short-term memory. The long-term memory is very slowly adaptive and stores listener-specific, context-independent localization cues such as spatial maps of ITDs, ILDs, and spectral shapes related to the filtering properties of the head and pinnae. In Plenge’s model, the short-term memory is quickly adaptive and represents context-dependent information about the sound source and the room such as visual information, motion trajectory, source intensity, and reverberation. He further concluded that externalization mainly relies on how well the currently accessible cues match the listener’s expectations based on short-term memory; only under conditions in which there is insufficient contextual information do matches with long-term memory representations become important. Many recent findings on sound externalization, reviewed in this paper, are consistent with this conceptual model.

Current quantitative approaches build on Plenge’s concept that signals providing inconsistencies with natural or stored information (or templates) tend to disrupt externalization. Figure 4 illustrates the basic structure of these template-based models. Hassager et al. (2016) proposed a model evaluating the deviation of spectral ILD profiles from those present in a reference signal. This model was motivated by their experiment using spectrally smoothed BRIRs (described earlier). They showed that spectral smoothing led to larger ILD deviations, and thus to predictions of more internalization, in line with the experimental results. This raises the possibility that some of the other monaural effects that have been reported may have interaural explanations. The bottom panels of Figure 3 illustrate this point for the spectral smoothing and spectral flattening methods of Hassager et al. (2016) and Baumgartner et al. (2017). Because these methods are applied independently to the two ears, what results is a change in the ILD spectrum relative to the unprocessed case. Recently, Baumgartner and Majdak (2020) applied the ILD deviation metric and other metrics to a larger number of experiments (Baumgartner et al., 2017; Boyd et al., 2012; Hartmann & Wittenberg, 1996; Hassager et al., 2016). The metrics included monaural and interaural spectral profiles, monaural and interaural spectral variance, IC, and inconsistencies between ITD and ILD. They concluded that a model that jointly assesses deviations in both monaural and interaural spectral profiles from the natural/unprocessed case could most reliably account for externalization ratings obtained under anechoic (or near-anechoic) conditions. Moreover, their simulations suggest that externalization ratings across experiments are based on a static—not dynamic—combination of those cues. One implication of this is that adaptation effects would reflect updated templates rather than changes in the weighting of different cues.



**Figure 4.** General structure of template-based externalization models that calculate the deviation of acoustical cues from those of a natural, well-externalized, reference stimulus. Models focused on the direct sound use long-term excitation patterns and involve comparisons of monaural cues at each ear and/or interaural cues (e.g., Baumgartner & Majdak, 2020; Hassager et al. 2016). Models focused on reverberation use short-term excitation patterns and involve only interaural-cue comparisons (e.g., Li et al. 2018).

A related approach has also been applied to externalization data collected under reverberant conditions. Building on the observations made by Catic et al. (2013, 2015), Li et al. (2018) formalized a quantitative model based on temporal fluctuations in the ILD and IC. This model consists of auditory peripheral filtering and the calculation of binaural cues within 20-ms segments. The standard deviation across segments, averaged across frequency channels, produces a single metric that can be mapped to externalization scores. The model simulations based on either ILD or IC provided reasonable predictions of the average externalization ratings from their experiment in which the BRIRs were truncated at the ipsilateral or contralateral ear or both (described earlier).

## Externalization and Hearing Technologies

### A/VR Systems

One of the major goals of A/VR systems is to create virtual environments that are perceived as realistic. A/VR has applications in communications, entertainment, gaming, and architecture (e.g., Jeon & Jo, 2019; Vorländer, 2020). While historically the emphasis in A/VR systems has been on the visual display, the audio component is gaining more and more attention. Moreover, acoustic A/VR systems are now in widespread use for a variety of applications (Xie, 2013). In many A/VR systems, however, technical sacrifices degrade the quality of sound reproduction. The achievable quality has been described in a variety of different ways. For example, Carlile (1996, p. 226) writes about “creating realistic environments that give users a feeling of being present in the created location.” Lindau and Weinzierl (2012, p. 804) used “plausibility” to describe “the perceived agreement with the listener’s expectation towards a corresponding real acoustic event.” These kinds of descriptions capture a complex combination of attributes that naturally includes sound externalization.

A/VR systems can rely on loudspeaker-based sound reproduction or on personal sound delivered by headphones or earphones. Most loudspeaker-based A/VR systems consist of many stationary loudspeakers, each of which represents a sound source that is stationary in space and thus usually can be well externalized (e.g., Brimijoin et al., 2013; Seeber et al., 2010). Wavefield synthesis and Ambisonics are sound field reproduction approaches that can be used for spatially continuous representations (Spors et al., 2013), and methods have been proposed to improve perceived externalization in such systems (Miller & Rafaely, 2019). Other loudspeaker-based A/VR systems aim to provide a dedicated binaural signal to the listener by canceling the cross talk in the acoustic paths between the individual loudspeakers and the listener’s individual ears (e.g., Akeroyd et al., 2007). However, even small deviations in this process yield artifacts that distort the monaural spectral cues, clearly degrading sound-localization performance (Majdak et al., 2013) and likely also disturbing externalization.

A/VR systems relying on personal sound devices can more accurately deliver binaural signals to the listener’s ears. When done correctly, static virtual sources created using listener-specific HRTFs cannot be distinguished from real sources (Hartmann & Wittenberg, 1996; Langendijk & Bronkhorst, 2000), implying that externalization can be captured along with many other characteristics. However, maintaining externalization under dynamic conditions brings extra challenges. To avoid a breakdown of externalization, virtual sounds must be consistent with expectations built upon proprioception, which includes appropriate changes in location with no perceivable latency (e.g., Brimijoin et al., 2013; Brungart et al., 2005). Moreover, room acoustics must be simulated appropriately (Gil-Carvajal et al., 2016; Savioja et al., 1999), and a particular challenge for AR systems is how to provide reverberant sounds that are perceptually consistent with the real environment.

Despite the various challenges, it is now possible to achieve well-externalized sound reproduction in



laboratory-based A/VR systems, and recent consumer products are available that incorporate externalized audio (e.g., Microsoft HoloLens, Facebook Reality Labs' Oculus products, Sony 360 Reality Audio, Genelec Aural ID). Another promising application of acoustic VR systems is in the assessment of hearing aids and cochlear implants, where it may be useful for the listener to be immersed in the kinds of environment they are likely to encounter in everyday life (e.g., Kressner et al., 2018; Oreinos & Buchholz, 2015; Pausch et al., 2018).

### Hearing Aids

A topic that has received surprisingly little attention in the literature is how hearing aids, which are worn on the ears and clearly interrupt the natural sound path, affect sound externalization. Anecdotal reports of internalized sound images are common in hearing-aid wearers, especially when first fit, although there is a scarcity of empirical data. Externalization is addressed in the commonly used Speech, Spatial and Qualities of Hearing Scale (SSQ; Gatehouse & Noble, 2004), a questionnaire used to assess real-life impairment and disability in listeners with hearing loss. One item reads, "Do the sounds of things you are able to hear seem to be inside your head rather than out there in the world?" and respondents give ratings on a continuous scale from 0 (*Inside my head*) to 10 (*Out there*). In a study that administered the SSQ to a large number of hearing-aid users (Noble & Gatehouse, 2006), a disruptive effect of hearing aids on externalization was confirmed, with lower scores for unilateral fittings relative to unaided, and for bilateral fittings relative to unilateral fittings.

The reason why hearing aids can cause internalization is not entirely clear. Because many hearing-aid earmolds result in a (fully or partially) blocked ear canal, one possibility is that internalization is related to the occlusion effect. The occlusion effect most often refers to a change in the perception of one's own voice and other internally generated sounds with blocked ear canals (Dillon, 2012). It is possible, though, that external sounds are also affected and that the occlusion effect may reflect a general change in one's sense of connection to the environment. Indeed the developers of the SSQ note that the item on internalization was included precisely because hearing-aid earmolds cause occlusion. Open earmolds have been studied for their potential beneficial effects on localization (e.g., Byrne et al., 1996, 1998), but there are no rigorous studies of externalization for different earmold fittings. A pertinent footnote can be found in Byrne et al. (1998, p. 71):

Our personal observations from wearing closed earmolds and earplugs are that sounds often do not seem

"out there" in their true locations. Furthermore, some of the hearing-impaired listeners in previous experiments have commented on greater externalization of sounds when changing from closed to open earmolds. However, the issue of externalization has not been studied and warrants investigation.

Somewhat surprisingly, a recent review of open versus closed fittings did not mention sound externalization (Winkler et al., 2016).

Another possibility is that hearing aids disrupt externalization by distorting the binaural cues received by the wearer. There are a number of ways in which hearing-aid processing may distort ITDs and ILDs (Brown et al., 2016; Dillon, 2012). The reason why such distortions would compromise externalization is not clear, although unnatural combinations of ITD and ILD may be a factor. Wiggins and Seeber (2012) provided evidence for this idea using the example of fast-acting dynamic-range compression. In their study, normally hearing listeners rated the perceived externalization of various anechoic stimuli using a questionnaire item similar to the SSQ item described earlier. Externalization ratings were reduced when compression was active (relative to those for unprocessed stimuli), an effect that the authors attributed to a reduction in ILDs and the resulting conflict with the undistorted ITD cues. It is not clear if similar effects would be observed for slow-acting compression or for listeners with hearing loss. Hassager, Wiinberg, et al. (2017) showed that the primary effects of fast-acting compression on reverberant stimuli were an increased level of the reverberation relative to the direct sound and a reduced IC. They argued that these physical distortions drove various perceptual distortions (including split/broad images and some instances of internalization) that were experienced by their normally hearing and hearing-impaired listeners. It has also been noted that hearing-aid compression can have particularly complex and unpredictable effects on binaural cues when the listener's head is moving (Brimijoin et al., 2017); in this case, a disruption of self-motion cues may further compromise externalization with hearing aids.

Externalization may also be affected by the placement of the microphone in some hearing aids. For example, in behind-the-ear hearing aids, the microphone is located above the pinna and cannot capture natural pinna-related spectral information. As discussed earlier, this may compromise externalization, and some evidence for this was provided by Boyd et al. (2012) for normally hearing listeners. Microphone placement is also one of the possible explanations for the spatial distortions (including poor externalization) reported by Cubick et al. (2018) for normally hearing listeners fitted with behind-the-ear devices. On the other hand, the impact

of spectral cue distortions may be tempered for hearing-impaired listeners who often have limited audibility in the relevant high-frequency regions. Also, many negative effects of hearing aids (including distortions of spatial perception and externalization) may decline after a period of exposure as listeners calibrate to the altered cues or simply acclimatize to the acoustical changes. This general point has been made previously as it also applies to the internalization associated with headphones (Durlach et al., 1992; Plenge, 1974) and nonindividualized HRTFs (Mendonça et al., 2013).

Finally, it is worth mentioning several recent studies that have explored methods for adding realistic externalization cues (via HRTFs and room reverberation) to hearing-aid signals that would otherwise be internalized (Courtois et al., 2019; Kates et al., 2018, 2019). This line of work demonstrates the importance that hearing-aid wearers and hearing-aid manufacturers place on natural, externalized sound images.

## Conclusions and Areas for Further Research

The main goal of this article was to provide an up-to-date review of the topic of sound externalization. We intentionally focused on recent studies that have provided new insights since previous reviews (Blauert, 1997; Durlach et al., 1992). These recent studies continue to provide support for the long-held idea that sound externalization relies on the signals received at the ears meeting the listener's expectations based on their past experiences and the current context. These studies also confirm the generally positive impact of reverberation on sound externalization and suggest that the binaural properties of the reverberation are particularly important. Other results suggest that pinna-related spectral cues are important for externalization, although their individualization is less critical in reverberant environments. Recent studies have also considered the contribution of visual information and of dynamic cues related to self-motion, concluding that externalization can be compromised if these cues are implausible or inconsistent with other cues.

Another goal of the article was to briefly review the current state of A/VR systems with respect to sound externalization, as these systems depend on providing convincing sound images as one factor in creating a realistic and immersive experience. The research covered in this review suggests that reverberation and plausible self-motion cues are key considerations for a compelling sense of externalization, while personalized HRTFs and congruent visual information may strengthen the illusion in some circumstances. Of course, the interaction of these factors is complex, and it is simply not possible

to provide fixed guidelines for determining their priority. However, research in this area is expected to continue at a fast pace as the challenges associated with wearable systems for consumer products are met.

A final goal of the article was to bring to attention the issue of sound externalization in hearing-aid wearers. We reviewed the evidence suggesting that hearing aids can disrupt sound externalization and explored the potential explanations for this disruption. We believe this is an important area for future research, which has implications not only for hearing aids but also for other devices that interrupt the natural sound path (hearing protectors, earphones, hearables, etc.).

Many open questions remain, for which further investigations would be useful. For example, the relationship between externalization and distance perception remains difficult to clarify given that current experimental approaches confound them. Also, recent studies using audiovisual environments have raised intriguing questions about the influence of context on sound externalization. A topic that was not addressed explicitly but was apparent from reviewing the literature is that there can be extremely large individual differences in externalization percepts (e.g., Baumgartner et al., 2017; Catic et al., 2015; Hartmann & Wittenberg, 1996; Kim & Choi, 2005; Udesen et al., 2015). It would be of great interest to understand the individual factors driving these differences. Another important area for future research concerns how temporal factors such as adaptation affect sound externalization; more insight on that topic would be particularly informative as it relates to training and acclimatization with hearing technologies. Finally, while quantitative approaches for explaining the degree of externalization have recently emerged, it seems clear that more empirical data will be needed to move toward a predictive model that takes into account acoustical parameters of both the direct and reverberant sounds, information from other modalities, individual factors, and adaptation.

## Acknowledgments

The authors are grateful to Owen Brimijoin and Jim Kates for providing raw data for Figure 2.

## Declaration of Conflicting Interests



The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors were supported by NIH/NIDCD grant number DC015760 (V. B.), Austrian Science Fund (FWF) grant

number I4294-B (R. B.), and grant ASH (PHC Danube 2020, 45268RE: M. L., APVV DS-FR-19-0025: N. K., WTZ MULT 07/2020: P. M.). M. L. is part of the LabEx CeLyA (ANR-10-LABX-0060/ANR-16-IDEX-0005). N. K. was also supported by VEGA 1/0355/20.

### ORCID iDs

Virginia Best  <https://orcid.org/0000-0002-5535-5736>  
Piotr Majdak  <https://orcid.org/0000-0003-1511-6164>

### References

- Akeroyd, M. A., Chambers, J., Bullock, D., Palmer, A. R., Summerfield, Q., Nelson, P. A., & Gatehouse, S. (2007). The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics. *Journal of the Acoustical Society of America*, *121*(2), 1056–1069. <https://doi.org/10.1121/1.2404625>
- Baumgartner, R., & Majdak, P. (2020). Decision making in auditory externalization perception. *bioRxiv*. <https://doi.org/10.1101/2020.04.30.068817>
- 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 of the United States of America*, *114*(36), 9743–9748. <https://doi.org/10.1073/pnas.1703247114>
- Begault, D. R., & Wenzel, E. M. (1993). Headphone localization of speech. *Human Factors*, *35*(2), 361–376. <https://doi.org/10.1177/001872089303500210>
- Begault, D. R., Wenzel, E. M., & Anderson, M. R. (2001). 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*(10), 904–916.
- Bidart, A., & Lavandier, M. (2016). Room-induced cues for the perception of virtual auditory distance with stimuli equalized in level. *Acta Acustica united with Acustica*, *102*, 159–169. <https://doi.org/10.3813/AAA.918933>
- Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization* (Rev. ed.). 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. *Journal of the Acoustical Society of America*, *131*(3), EL268–EL274. <https://doi.org/10.1121/1.3687015>
- Brimijoin, W. O., & Akeroyd, M. A. (2012). The role of head movements and signal spectrum in an auditory front/back illusion. *i-Perception*, *3*(3), 179–182. <https://doi.org/10.1068/i7173sas>
- 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>
- Brimijoin, W. O., McLaren, A. I., & Naylor, G. M. (2017). Hearing impairment, hearing aids, and cues for self motion. *Proceedings of Meetings on Acoustics*, *28*, 050003. <https://doi.org/10.1121/2.0000359>
- Brown, A. D., Rodriguez, F. A., Portnuff, C. D., Goupell, M. J., & Tollin, D. J. (2016). Time-varying distortions of binaural information by bilateral hearing aids: Effects of nonlinear frequency compression. *Trends in Hearing*, *20*. <https://doi.org/10.1177/2331216516668303>
- Brungart, D. S., Simpson, B. D., & Kordik, A. J. (2005). The detectability of headtracker latency in virtual audio displays. *Proceedings of the 11th Meeting of the International Conference on Auditory Display*, 37–42.
- Byrne, D., Noble, W., & Glauerdt, B. (1996). Effects of earmold type on ability to locate sounds when wearing hearing aids. *Ear and Hearing*, *17*(3), 218–228. <https://doi.org/10.1097/00003446-199606000-00005>
- Byrne, D., Sinclair, S., & Noble, W. (1998). Open earmold fittings for improving aided auditory localization for sensorineural hearing losses with good high-frequency hearing. *Ear and Hearing*, *19*(1), 62–71. <https://doi.org/10.1097/00003446-199802000-00004>
- Callan, A., Callan, D. E., & Ando, H. (2013). Neural correlates of sound externalization. *NeuroImage*, *66*, 22–27. <https://doi.org/10.1016/j.neuroimage.2012.10.057>
- Carlile, S. (1996). *Virtual auditory space: Generation and applications*. Landes Bioscience.
- Carlile, S., Ciccarelli, G., Cockburn, J., Diedesch, A. C., Finnegan, M. K., Hafter, E., Henin, S., Kalluri, S., Kell, A. J. E., Ozmeral, E. J., Roark, C. L., & Sagers, J. E. (2017). Listening into 2030 workshop: An experiment in envisioning the future of hearing and communication science. *Trends in Hearing*, *21*, 2331216517737684. <https://doi.org/10.1177/2331216517737684>
- Catic, J., Santurette, S., Buchholz, J. M., Gran, F., & Dau, T. (2013). The effect of interaural-level-difference fluctuations on the externalization of sound. *Journal of the Acoustical Society of America*, *134*(2), 1232–1241. <https://doi.org/10.1121/1.4812264>
- Catic, J., Santurette, S., & Dau, T. (2015). The role of reverberation-related binaural cues in the externalization of speech. *Journal of the Acoustical Society of America*, *138*(2), 1154–1167. <https://doi.org/10.1121/1.4928132>
- Courtois, G., Grimaldi, V., Lissek, H., Estoppey, P., & Georganti, E. (2019). Perception of auditory distance in normal-hearing and moderate-to-profound hearing-impaired listeners. *Trends in Hearing*, *23*. <https://doi.org/10.1177/2331216519887615>
- Cubick, J., Buchholz, J. M., Best, V., Lavandier, M., & Dau, T. (2018). Listening through hearing aids affects spatial perception and speech intelligibility in normal-hearing listeners. *Journal of the Acoustical Society of America*, *144*(5), 2896–2905. <https://doi.org/10.1121/1.5078582>
- Dillon, H. (2012). *Hearing aids* (2nd ed.). Boomerang Press.
- Durlach, N. I., Rigopoulos, A., Pang, X. D., Woods, W. S., Kulkarni, A., Colburn, H. S., & Wenzel, E. M. (1992). On the externalization of auditory images. *Presence: Teleoperators & Virtual Environments*, *1*(2), 251–257. <https://doi.org/10.1162/pres.1992.1.2.251>
- Gatehouse, S., & Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *International Journal of Audiology*, *43*, 85–99. <https://doi.org/10.1080/14992020400050014>

- 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>
- Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1–2), 103–138. [https://doi.org/10.1016/0378-5955\(90\)90170-t](https://doi.org/10.1016/0378-5955(90)90170-t)
- Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *British Medical Journal*, *331*(7524), 1064–1065. <https://doi.org/10.1136/bmj.38636.593461.68>
- Hartmann, W. M., & Wittenberg, A. (1996). On the externalization of sound images. *Journal of the Acoustical Society of America*, *99*(6), 3678–3688. <https://doi.org/10.1121/1.414965>
- 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. *Journal of the Acoustical Society of America*, *139*(5), 2992–3000. <https://doi.org/10.1121/1.4950847>
- Hassager, H. G., May, T., Wiinberg, A., & Dau, T. (2017). Preserving spatial perception in rooms using direct-sound driven dynamic range compression. *Journal of the Acoustical Society of America*, *141*(6), 4556–4566. <https://doi.org/10.1121/1.4984040>
- Hassager, H. G., Wiinberg, A., & Dau, T. (2017). Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment. *Journal of the Acoustical Society of America*, *141*(4), 2556–2568. <https://doi.org/10.1121/1.4979783>
- Hendrickx, E., Stitt, P., Messonnier, J.-C., Lyzwa, J.-M., Katz, B. F. G., & de Boishéraud, C. (2017a). 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., & de Boishéraud, C. (2017b). Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis. *Journal of the Acoustical Society of America*, *141*(3), 2011–2023. <https://doi.org/10.1121/1.4978612>
- Hunter, M. D., Griffiths, T. D., Farrow, T. F. D., Zheng, Y., Wilkinson, I. D., Hegde, N., Woods, W., Spence, S. A., & Woodruff, P. W. R. (2003). A neural basis for the perception of voices in external auditory space. *Brain*, *126*(Pt 1), 161–169. <https://doi.org/10.1093/brain/awg015>
- Jeffress, L. A., & Taylor, R. W. (1961). Lateralization vs. localization. *Journal of the Acoustical Society of America*, *33*, 482–483. <https://doi.org/10.1121/1.1908697>
- Jeon, J. Y., & Jo, H. I. (2019). Three-dimensional virtual reality-based subjective evaluation of road traffic noise heard in urban high-rise residential buildings. *Building and Environment*, *148*, 468–477. <https://doi.org/10.1016/j.buildenv.2018.11.004>
- Jiang, Z., Sang, J., Zheng, C., & Li, X. (2020). The effect of pinna filtering in binaural transfer functions on externalization in a reverberant environment. *Applied Acoustics*, *164*, 107257. <https://doi.org/10.1016/j.apacoust.2020.107257>
- Jones, B., & Kabanoff, B. (1975). Eye movements in auditory space perception. *Perception & Psychophysics*, *17*, 241–245. <https://doi.org/10.3758/BF03203206>
- Kates, J. M., Arehart, K. H., & Harvey, L. O. (2019). Integrating a remote microphone with hearing-aid processing. *Journal of the Acoustical Society of America*, *145*(6), 3551–3566. <https://doi.org/10.1121/1.5111339>
- 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. *Journal of the Acoustical Society of America*, *143*(5), 2666–2677. <https://doi.org/10.1121/1.5032326>
- Kim, S. M., & Choi, W. (2005). On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach. *Journal of the Acoustical Society of America*, *117*(6), 3657–3665. <https://doi.org/10.1121/1.1921548>
- 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. <https://doi.org/10.17743/jaes.2016.0072>
- Kolarik, A. J., Moore, B. C. J., Zahorik, P., Cirstea, S., & Pardhan, S. (2016). Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, & Psychophysics*, *78*(2), 373–395. <https://doi.org/10.3758/s13414-015-1015-1>
- Kopčo, N., Doreswamy, K. K., Huang, S., Rossi, S., & Ahveninen, J. (2020). Cortical auditory distance representation based on direct-to-reverberant energy ratio. *NeuroImage*, *208*, 116436. <https://doi.org/10.1016/j.neuroimage.2019.116436>
- Kopčo, N., Huang, S., Belliveau, J. W., Raij, T., Tengshe, C., & Ahveninen, J. (2012). Neuronal representations of distance in human auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(27), 11019–11024. <https://doi.org/10.1073/pnas.1119496109>
- Kressner, A. A., Westermann, A., & Buchholz, J. M. (2018). The impact of reverberation on speech intelligibility in cochlear implant recipients. *Journal of the Acoustical Society of America*, *144*(2), 1113–1122. <https://doi.org/10.1121/1.5051640>
- Kulkarni, A., & Colburn, H. S. (1998). Role of spectral detail in sound-source localization. *Nature*, *396*, 747–749. <https://doi.org/10.1038/25526>
- Langendijk, E. H., & Bronkhorst, A. W. (2000). Fidelity of three-dimensional sound reproduction using a virtual auditory display. *Journal of the Acoustical Society of America*, *107*(1), 528–537. <https://doi.org/10.1121/1.428321>
- Leclère, T., Lavandier, M., & Perrin, F. (2019). On the externalization of sound sources with headphones without reference to a real source. *Journal of the Acoustical Society of America*, *146*(4), 2309–2320. <https://doi.org/10.1121/1.5128325>

- Levy, E. T., & Butler, R. A. (1978). Stimulus factors which influence the perceived externalization of sound presented through headphones. *The Journal of Auditory Research*, 18, 41–50.
- 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. *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(460), 1–16. <https://doi.org/10.3390/app9030460>
- Lindau, A., & Weinzierl, S. (2012). Assessing the plausibility of virtual acoustic environments. *Acta Acustica united with Acustica*, 98(5), 804–810. <https://doi.org/10.3813/AAA.918562>
- Loomis, J. M. (1992). Distal attribution and presence. *Presence: Teleoperators & Virtual Environments*, 1(1), 113–119. <https://doi.org/10.1162/pres.1992.1.1.113>
- Loomis, J. M., Hebert, C., & Cicinelli, J. G. (1990). Active localization of virtual sounds. *Journal of the Acoustical Society of America*, 88(4), 1757–1764. <https://doi.org/10.1121/1.400250>
- 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.
- Majdak, P., Goupell, M. J., & Laback, B. (2010). 3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training. *Attention, Perception, & Psychophysics*, 72, 454–469. <https://doi.org/10.3758/APP.72.2.454>
- Majdak, P., Masiero, B., & Fels, J. (2013). Sound localization in individualized and non-individualized crosstalk cancellation systems. *Journal of the Acoustical Society of America*, 133(4), 2055–2068. <https://doi.org/10.1121/1.4792355>
- 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>
- Miller, E., & Rafaely, B. (2019). The role of direct sound spherical harmonics representation in externalization using binaural reproduction. *Applied Acoustics*, 148, 40–45. <https://doi.org/10.1016/j.apacoust.2018.12.011>
- Noble, W., & Gatehouse, S. (2006). Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the Speech, Spatial, and Qualities of Hearing Scale (SSQ). *International Journal of Audiology*, 45(3), 172–181. <https://doi.org/10.1080/14992020500376933>
- Oreinos, C., & Buchholz, J. M. (2015). Objective analysis of ambisonics for hearing aid applications: Effect of listener's head, room reverberation, and directional microphones. *Journal of the Acoustical Society of America*, 137(6), 3447–3465. <https://doi.org/10.1121/1.4919330>
- Pausch, F., Aspöck, L., Vorländer, M., & Fels, J. (2018). An extended binaural real-time auralization system with an interface to research hearing aids for experiments on subjects with hearing loss. *Trends in Hearing*, 22. <https://doi.org/10.1177/2331216518800871>
- Plenge, G. (1972). On the problem of “In Head Localization”. *Acta Acustica united with Acustica*, 26(5), 241–252.
- Plenge, G. (1974). On the differences between localization and lateralization. *Journal of the Acoustical Society of America*, 56(3), 944–951. <https://doi.org/10.1121/1.1903353>
- Robinson, P., & Xiang, N. (2013). Design, construction, and evaluation of a 1:8 scale model binaural manikin. *Journal of the Acoustical Society of America*, 133(3), EL162–EL167. <https://doi.org/10.1121/1.4789876>
- Savioja, L., Huopaniemi, J., Lokki, T., & Väänänen, R. (1999). Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society*, 47(9), 675–705.
- Seeber, B., Kerber, S., & Hafter, E. (2010). A system to simulate and reproduce audio-visual environments for spatial hearing research. *Hearing Research*, 260(1–2), 1–10. <https://doi.org/10.1016/j.heares.2009.11.004>
- Shinn-Cunningham, B. G., Santarelli, S., & Kopco, N. (2000). Tori of confusion: Binaural localization cues for sources within reach of a listener. *Journal of the Acoustical Society of America*, 107(3), 1627–1636. <https://doi.org/10.1121/1.428447>
- Simon, L. S., Zacharov, N., & Katz, B. F. (2016). Perceptual attributes for the comparison of head-related transfer functions. *Journal of the Acoustical Society of America*, 140(5), 3623–3632. <https://doi.org/10.1121/1.4966115>
- Spors, S., Wierstorf, H., Raake, A., Melchior, F., Frank, M., & Zotter, F. (2013). Spatial sound with loudspeakers and its perception: A review of the current state. *Proceedings of the IEEE*, 101(9), 1920–1938. <https://doi.org/10.1109/JPROC.2013.2264784>
- Toole, F. E. (1970). In-head localization of acoustic images. *Journal of the Acoustical Society of America*, 48, 943–949. <https://doi.org/10.1121/1.1912233>
- Udesen, J., Piechowiak, T., & Gran, F. (2015). The effect of vision on psychoacoustic testing with headphone-based virtual sound. *Journal of the Audio Engineering Society*, 63, 552–561. <https://doi.org/10.17743/jaes.2015.0061>
- Vorländer, M. (2020). Are virtual sounds real? *Acoustics Today*, 16(1), 46–54. <https://doi.org/10.1121/AT.2020.16.1.46>
- Wallach, H. (1940). The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, 27, 339–368. <https://doi.org/10.1037/h0054629>
- Weber, E. H. (1849). Ueber die Umstände durch welche man geleitet wird manche Empfindungen auf äussere Objecte zu beziehen *Berichte Ueber die Verhandlungen der Koeniglichen Saechsischen Gesellschaft der Wissenschaften (Mathematisch-Physische Klasse)*, 2, 226–237.
- Werner, S., Klein, F., Mayenfels, T., & Brandenburg, K. (2016). A summary on acoustic room divergence and its effect on externalization of auditory events. *8th International Conference on Quality of Multimedia Experience*, 1–6. <https://doi.org/10.1109/QoMEX.2016.7498973>
- Wiggins, I. M., & Seeber, B. U. (2012). Effects of dynamic-range compression on the spatial attributes of sounds in

- normal-hearing listeners. *Ear and Hearing*, 33(3), 399–410. <https://doi.org/10.1097/AUD.0b013e31823d78fd>
- Wightman, F. L., & Kistler, D. J. (1999). Resolution of front-back ambiguity in spatial hearing by listener and source movement. *Journal of the Acoustical Society of America*, 105(5), 2841–2853. <https://doi.org/10.1121/1.426899>
- Winkler, A., Latzel, M., & Holube, I. (2016). Open versus closed hearing-aid fittings: A literature review of both fitting approaches. *Trends in Hearing*, 20. <https://doi.org/10.1177/2331216516631741>
- Xie, B. (2013). *Head-related transfer function and virtual auditory display* (2nd ed.). J. Ross Publishing.
- Yuan, Y., Xie, L., Fu, Z.-H., Xu, M., & Cong, Q. (2017). Sound image externalization for headphone based real-time 3D audio. *Frontiers of Computer Science*, 11(3), 419–428. <https://doi.org/10.1007/s11704-016-6182-2>
- Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica*, 91, 409–420.
- Zeng, F.-G. (2016). Goodbye Google Glass, hello smart ear-phones. *The Hearing Journal*, 69(12), 6. <https://doi.org/10.1097/01.HJ.0000511124.64048.c1>