On the externalization of sound sources with headphones without reference to a real source

Thibaud Leclère, Mathieu Lavandier, Fabien Perrin

On the externalization of sound sources with headphones without reference to a real source

Thibaud Leclère,[1,a)] Mathieu Lavandier,[1] and Fabien Perrin[2]
[1]*Univ Lyon, ENTPE, Laboratoire Génie Civil et Bâtiment, Rue Maurice Audin, F-69518 Vaulx-en-Velin Cedex, France*
[2]*Auditory Cognition and Psychoacoustics Team, Lyon Neurosciences Research Center, UCBL-CNRS-INSERM 5292, Inserm U1028, Lyon, France*

(Received 2 May 2019; revised 6 September 2019; accepted 16 September 2019; published online 10 October 2019)

Sounds presented over headphones are generally perceived as internalized, i.e., originating from a source inside the head. Prior filtering by binaural room impulse responses (BRIRs) can create externalized sources. Previous studies concluded that these BRIRs need to be listener-specific to produce good externalization; however, listeners were generally facing a loudspeaker and asked to rate externalization relative to that loudspeaker, meaning that the source had to be perceived outside the head and also at the right distance. The present study investigated externalization when there is no visual source to match. Overall, lateral sources were perceived as more externalized than frontal sources. Experiment 1 showed that the perceived externalization obtained with non-individualized BRIRs measured in three different rooms was similar to that obtained with a state-of-the-art simulation using individualized BRIRs. Experiment 2 indicated that when there is no real source spectrum to match, headphone equalization does not improve externalization. Experiment 3 further showed that reverberation improved externalization only when it introduced interaural differences. Correlation analyses finally showed a close correspondence between perceived externalization and binaural cues (especially interaural coherence).
https://doi.org/10.1121/1.5128325

[MD] Pages: 2309–2320

I. INTRODUCTION

Under natural listening conditions, sound sources are perceived as "externalized," meaning that they are heard as coming from outside the head and can clearly be identified as being part of the sound environment of the listener. Conversely, sounds presented using headphones are most often perceived inside the head ("internalized"). The main difference between these two listening modes is often attributed to the filtering of the incoming sound by the listener's head and torso characterized by the head-related transfer functions (HRTFs). When listening over headphones, the acoustic signal directly reaches the eardrum, discarding the natural filtering by the HRTFs occurring for sounds produced by a source external to the listener. Binaural synthesis involves the application of such filtering to the signals prior to delivering them through headphones in order to reproduce the acoustic signals received at the eardrum under natural listening conditions (Wightman and Kistler, 1989a,b). Since HRTFs depend on the morphology of each listener (e.g., shape of the pinnae, size of the head), binaural synthesis is most accurate when using HRTFs that match those of the listeners. However, measuring individualized HRTFs can be very time-consuming and cannot always be done in practice, so non-individualized HRTFs (e.g., measured on a manikin) are often used.

The aim of the present study was to evaluate how well externalized a sound source could be simulated with simple static binaural synthesis, i.e., without dynamic rendering, individual HRTFs, or headphone equalization, but also without any constraint on its perceived location in the environment outside the listener's head. In particular, there was no requirement to match the position of a real (even silent) source visible to the listeners, who were instructed to keep their eyes closed while listening to the stimulus.

In rooms, the filtering by the HRTFs is complemented by the transfer function of the room (one for each ear). The combination of these two functions is described in the time domain by binaural room impulse responses (BRIRs) measured at the ears of a listener (individualized) or manikin (non-individualized) for a source at a given position within the room. BRIRs can be used in binaural synthesis just as HRTFs in order to simulate reverberant environments. Reverberation was previously demonstrated to enhance perceived externalization (Kates *et al.*, 2018; Begault *et al.*, 2001). Catic *et al.* (2013) suggested that this enhancement could be due to the fact that reverberation induces temporal variations of interaural level differences (ILDs). They observed that, for stimuli containing energy above 1 kHz, externalization decreased when reducing the temporal fluctuations of ILDs. Also, listeners always indicated an internalized percept when listening monaurally. Catic *et al.* (2015) further interrogated the role played by binaural cues within the BRIR. While keeping the early part of the BRIR binaural, the late part was either removed or replaced by diotic

[a)]Electronic mail: tleclere@usal.es

reverberation. The temporal cutoff between the early and late parts was varied, leading to more or less access to the binaural cues. They identified that binaural cues in the early part of the BRIR were necessary to externalize. Externalization increased when increasing the early-binaural/late-diotic cutoff from 20 to about 80 ms (highlighting the importance of early reflections) and then plateaued until 500 ms (full binaural BRIR). This study also suggested that binaural cues from reflections are more important for externalization when the direct sound itself contains only weak binaural cues (e.g., for frontal sources) than when the direct sound contains larger interaural cues (e.g., for lateral sources). Li *et al.* (2018) investigated the influence of reverberation on externalization of lateral sources. By manipulating reverberation in each ear separately, they determined that reverberation received by the contralateral ear had more influence on externalization compared to the ipsilateral ear. Li *et al.* (2019) further studied the contributions of ipsilateral vs contralateral ears on externalization while varying the azimuth of the source. As a result, the contribution of the reverberation at the contralateral ear increased as the source moved to the side.

Externalization has been investigated by comparing headphone listening to loudspeaker listening (Kulkarni and Colburn, 1998; Hartmann and Wittenberg, 1996) or to virtual-loudspeaker listening (i.e., the loudspeaker was visible, remained silent during the experiment, and was simulated over headphones; Catic *et al.*, 2013; Catic *et al.*, 2015). In both cases, listeners were asked to rate externalization relative to a visible loudspeaker. For example, the scale used by Catic *et al.* (2013) consisted of four possible ratings: "(0) the sound is in my head; (1) the sound is closer to me; (2) the sound is closer to the loudspeaker; and (3) the sound is at the position of the loudspeaker." With such a scale, the simulation through headphones not only has to produce externalization with the source perceived outside the head, but this source also has to be perceived at the right distance (the one of the silent loudspeaker). This is generally assured with individualized HRTFs/BRIRs and headphone equalization. To better disentangle externalization from distance perception, the present study investigated externalization without reference to a visual source.

Head movements have been shown to influence source externalization (Hendrickx *et al.*, 2017; Brimijoin *et al.*, 2013), leading to the fact that dynamic binaural synthesis (changing BRIRs in real time according to head movements) is often considered as a state-of-the art method to reproduce externalized/realistic sound sources. However, since the present study was focused on the aspect of externalization due to the auditory stimulus only, it did not involve head tracking, and subjects were then instructed to minimize their head motion as much as they could. Visual cues could also favor externalization, as observed in the ventriloquism effect, where sound localization is highly biased towards a visual reference varying in both azimuth and elevation (Hendrickx *et al.*, 2015) or in distance (Zahorik, 2001). The visual impression of a room can also influence distance perception (Calcagno *et al.*, 2012), and thus potentially affect externalization. To limit any potential bias due to vision, externalization was evaluated here in the absence of visual reference (eyes closed in a listening booth).

A first aim of the present study was to investigate whether the use of non-individualized stimuli could be sufficient to externalize a sound source in the absence of any constraint on its perceived location outside the listener's head. In Experiment 1, the perceived degree of externalization obtained with non-individualized BRIRs (NI-BRIRs) was quantitatively compared to the externalization achieved by a state-of-the-art simulation with individualized BRIRs (I-BRIRs), which constituted our reference for good externalization in the absence of a visual source. In Experiment 2, the method of simulation with NI-BRIRs was further tested by evaluating the influences of the headphones and their equalization. The aim of Experiment 3 was to investigate the relative contributions of binaural hearing and reverberation on externalization. Finally, a correlation analysis was performed on all stimuli used in these experiments to identify simple acoustical correlates that could be used to predict externalization.

BRIRs measured in seven different rooms (one being anechoic) were tested across Experiments 1 to 3. The aim here was not to link perceived externalization to specific room parameters, a goal beyond the scope of the present study, but rather to introduce ecologically relevant variability in the stimuli, so that the results would not be only associated with a particular room/reverberant condition.

II. GENERAL METHODS

A. Stimuli

Four original signals were used: pink noise, music, speech and an "environmental" sound (bottles). The pink noise was a 1.26-s burst. The music signal was an excerpt of jazz music (McCoy Tyner, "Miss Bea," Best of Chesky Jazz, Vol. 2, Chesky CD:68, from 01:01 to 01:05, right channel). The speech signal was a 0.9-s anechoic recording of three French words ("Toute la nuit" meaning "All night long") spoken by a male. The "bottles" signal was a 1-s recording of clinging glass bottles and was only used in Experiment 3.

In a given condition, binaural stimuli were created by convolving the original signal described above with a corresponding BRIR (see Sec. II B). In Experiment 1, the frequency response of the headphones was compensated for by convolving the binaural signals with the inverse impulse response of the headphones measured on the listener (individualized conditions) or on a manikin (non-individualized conditions). In Experiment 2, only the non-individualized inverse filter from the manikin was used in a single experimental condition. No headphone equalization was used in Experiment 3. Finally, all stimuli were equalized in level such that the average of the root-mean-square (RMS) power of the left- and right-ear signals was set to the same RMS power as a diotic white noise delivered at 69 dB sound pressure level (SPL) (68 dB SPL in Experiment 3).

In all experiments, unprocessed control conditions were also tested, in which the original signals were presented diotically without any processing except level equalization (i.e., neither BRIR convolution nor headphone equalization).

B. BRIR measurements

For Experiments 1 and 2, I-BRIRs were recorded in three different rooms (classroom, meeting room and gym) at two azimuths (0° and 60°) and two distances (1 and 5 m) from the listener (see Fig. 1). The room dimensions were the following (with the notation Length × Width × Height): 9.4 m × 10.7 m × 2.8 m (Classroom), 7.7 m × 10.5 m × 2.8 (Meeting room), 33.7 m × 44.5 m × 10.5 m (Gym). Recordings were carried out by using the log sine sweep technique (Farina, 2007) with a sweep duration of 15 s and a frequency range from 20 Hz to 20 kHz. A sine sweep signal was generated, converted into analog signal through a soundcard (RME Fireface 800), passed through an amplifier (Power amplifier Brüel & Kjær 2716) and finally delivered to a loudspeaker (Tannoy System 8 NFM 2) located at the desired position in the room. The acoustic signal was then recorded by two miniature microphones (Knowles FG 23 329-P7) wrapped in foam ear tips and positioned at the entrance of each ear canal of the listener pointing outwards. Electric signals were sent back to the RME soundcard for digitalization and recording. BRIRs were then computed by convolving the recorded sweeps by their corresponding reversed version along the time axis using the overlap-add algorithm (Oppenheim and Schafer, 2014). Listeners sat on a chair during each measurement and were instructed to look straight ahead, keeping their head still. The same method was used to measure NI-BRIRs in the same rooms for the same source locations using a Cortex manikin MK1. Headphone transfer functions were measured using the same technique while listeners wore the headphones (Sennheiser HD 650) on top of the

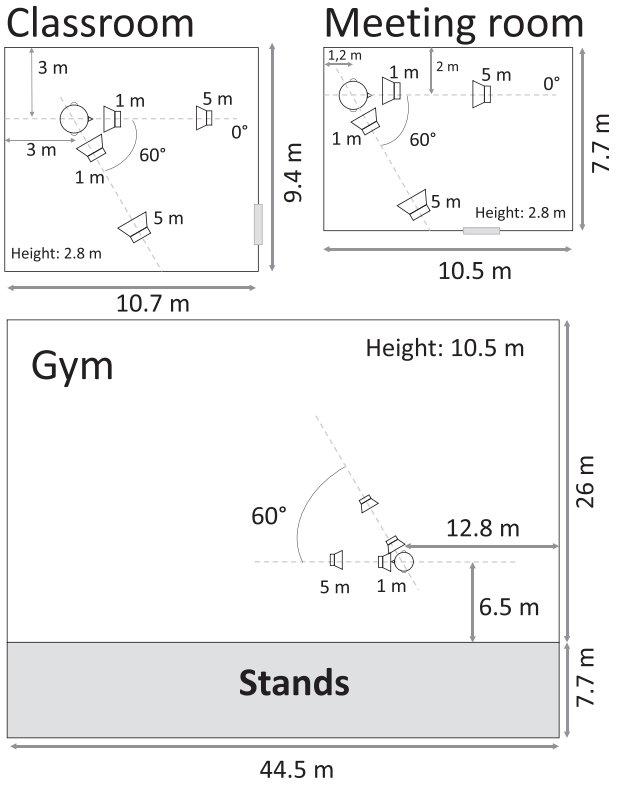

FIG. 1. Scaled layout of each room where BRIRs were measured, indicating the respective positions of the listener and of each loudspeaker in the room (the scale is different for each room).

microphones inserted in their ear canals. Ten consecutive measurements were made by removing and replacing the headphones between two measurements in order to take into account the variation due to headphones positioning on the head (Kulkarni and Colburn, 2000). The headphones' inverse impulse responses were then computed for each subject using linear phase filters[1] (Oppenheim and Schafer, 2014), and were finally averaged across the ten measurements, as suggested by Kulkarni and Colburn (2000). The same method was performed on the manikin to collect the non-individualized inverse response of the headphones.

For Experiment 3, stimuli were convolved with NI-BRIRs measured in different rooms by other research teams: anechoic, Salford 1, Salford 2, Surrey B, and Surrey C. Anechoic BRIRs and Surrey BRIRs (for both rooms B and C) were taken from the University of Surrey database.[2] Surrey B was a medium-small classroom with a reverberation time (RT60) of 0.47 s and a direct-to-reverberant ratio (DRR) of 5.31 dB, while Surrey C was a large cinema-style theatre (RT60 = 0.68 s; DRR = 8.82 dB; Hummersone *et al.*, 2010). BRIRs from Salford 1 and 2 were measured in the same room (6.6 m × 5.8 m × 2.8 m; RT60 = 0.27 s) at distances of 1 and 2 m, respectively. They were obtained from the University of Salford database[3] (for more details, see Satongar *et al.*, 2014). In each room, BRIRs corresponding to four azimuths were used: 0° (in front), 30°, 60°, and 90° (on the left). It should be noted that, in the Salford room, the impulse responses were obtained from only one loudspeaker position, by rotating the manikin on itself to achieve the desired azimuths. Conversely, impulse responses from the Surrey database were recorded by moving the loudspeaker around the manikin.

C. Listening test procedure

On each trial, participants followed the step-by-step instructions displayed on the graphical interface: (1) "Close your eyes[4] and play the sound. Once you are ready to make your judgment, open your eyes," (2) "how externalized was the sound you just listened to?" During playback, a cursor appeared at a random location on a continuous line with extremities labeled "perceived inside the head" (corresponding to 0% of externalization) and "completely externalized" (100% externalization). Listeners had to move the cursor with the computer mouse to enter their judgment and then validate it. The position of the cursor was then linearly mapped to the degree of perceived externalization in percent. After validation of the judgment, the following trial started to step 1 "Close your eyes and play the sound…" See the Appendix for detailed instructions given to the listeners. This method, already used by Li *et al.* (2018), was chosen over categorical scales (Hendrickx *et al.*, 2017; Catic *et al.*, 2013; Catic *et al.*, 2015; Hartmann and Wittenberg, 1996) or MUSHRA (multiple stimuli with hidden reference and anchor; Cubick *et al.*, 2015) where, in both cases, listeners need to process auditory stimuli and compare it to an audio or visual (or both) reference.

Listeners first began the experiment with a short practice session to get used to the task and the stimuli. Stimuli were presented in random order during the test and practice.

D. Apparatus

Stimuli were D/A converted and amplified using a Lynx TWO sound card and delivered through Sennheiser HD 650 headphones (Experiments 1–3) and Sennheiser IE4 earphones (Experiment 2). Listeners were seated in a double-walled soundproof booth. They interacted with the graphical interface displayed outside the booth window by using a keyboard (to play the stimuli) and a computer mouse (to report their answers).

E. Listeners

Eighteen listeners (nine women and nine men) participated in both the BRIR measurements and Experiment 1. Their age ranged between 22 and 34 years with an average of 26.8 years. All listeners presented hearing thresholds to pure tones inferior or equal to 25 dB hearing level (HL) at all frequencies (each octave from 125 Hz to 8 kHz, plus 6 kHz) on both ears. Eighteen listeners (12 women and 6 men) who did not take part in Experiment 1 with self-reported normal hearing participated in Experiment 2. Their age ranged between 19 and 24 years with an average of 20.9 years. Twenty-one listeners (9 women and 12 men) who self-reported normal hearing participated in Experiment 3. Their age ranged between 20 and 60 years with an average of 34.1 years. All listeners were paid for their participation and signed an informed consent before participating. Three listeners participated in two experiments (one in Experiments 2 and 3, and two in Experiments 1 and 3).

III. EXPERIMENT 1: INFLUENCE OF INDIVIDUALIZATION ON EXTERNALIZATION

A. Aim and design

In Experiment 1, the degree of perceived externalization obtained with non-individualized stimuli was quantitatively compared to the state-of-the-art externalization obtained with individualized stimuli involving headphone equalization to compensate for the frequency response of the headphones (tested here in the absence of a visual source reference).

The stimuli were created using the I-BRIRs and NI-BRIRs measured on the listeners and on a manikin, respectively, in three rooms (classroom, meeting room, and gym) for a source at two azimuths (0° and 60°) and two distances from the listener (1 and 5 m). The transfer functions of the headphones (measured on the listeners and manikin) were compensated for by inverse filtering. Unprocessed conditions were also tested where signals were presented diotically to the listener without any BRIR convolution or headphone equalization. Three signals were used: pink noise, music, and speech. This resulted in 75 stimuli presented in random order with three repetitions in a single block. In all experiments of this study, externalization scores of each subject were first averaged across repetitions prior to any analysis.

For the 15-trial practice session of Experiment 1, all three types of signals were presented diotically and convolved with four non-individualized BRIRs not used for the rest of the experiment. These BRIRs were taken from the University of Surrey database (Hummersone *et al.*, 2010), measured in an anechoic room and a reverberant room (Room C) at two azimuths (0° and 60°).

B. Results

The mean degrees of perceived externalization measured in Experiment 1 are plotted in Fig. 2. Every stimulus convolved with a BRIR was perceived more externalized than unprocessed stimuli. Results also indicate that using NI- or I-BRIRs led to almost similar externalization scores despite an overall slightly higher externalization with I-BRIRs (53.4% vs 51.2%). Lateral sources were rated as more externalized than frontal sources. Increasing source-listener distance increased externalization for frontal sources. This effect was generally not observed for lateral sources.

A repeated-measures analysis of variance (ANOVA) with five within-subjects factors (individualization, azimuth, distance, room, signal type) was performed on these results (omitting the unprocessed conditions). Main effects were found significant for: individualization [$F(1,17) = 8.87$; $p < 0.01$], azimuth [$F(1,17) = 28.09$; $p < 0.001$], distance [$F(1,17) = 6.94$; $p = 0.02$], and type of signal [$F(2,34) = 3.30$; $p = 0.048$]. Concerning this latest factor, a *post hoc* HSD Tukey analysis indicated that music was perceived more externalized (57%) than speech (48%; $p = 0.046$). Two significant two-way interactions were also observed: azimuth × distance [$F(1,17) = 6.3393$; $p = 0.022$], individualization × room [$F(2,34) = 6.59$; $p < 0.01$]. Finally, three-way interactions were found significant: azimuth × distance × room [$F(2,34) = 7.95$; $p = 0.001$], azimuth × room × signal type [$F(4,68) = 3.5$; $p = 0.01$].

Simple effect analyses with Bonferroni corrections were further performed on each significant interaction. For the interaction azimuth × distance, sources with a lateral azimuth (60°) were evaluated significantly more externalized ($p < 0.001$) than sources located at 0° for the two distances. Similarly, the 5-m distance condition was perceived more externalized compared to the 1-m distance condition at the two azimuths ($p < 0.05$).

The improvement of externalization by increasing the azimuth was slightly greater when the distance was 1 m than when it was 5 m (29 against 26 percentage points, respectively). For the interaction individualization × room, I-BRIRs led to a significantly higher externalization than NI-BRIRs only for the classroom and the gym ($p < 0.05$). Concerning the influence of the room, only the sources in the classroom were perceived significantly more externalized than the sources in the gym in the case of I-BRIRs ($p = 0.04$). Concerning the interaction azimuth × distance × room, the simple effect analysis first indicated that the lateral source led to significantly higher externalization ratings compared to the frontal source for all levels of source distance and room ($p < 0.001$). Second, far sources led to significantly higher ratings compared to close sources in the meeting room and the classroom for the frontal sources ($p < 0.025$), as well as in the gym for the lateral sources ($p = 0.005$). Third, perceived externalization was higher in the classroom than in the gym only for the far frontal source. For the interaction azimuth × room × signal type, externalization ratings for the lateral source were significantly higher than

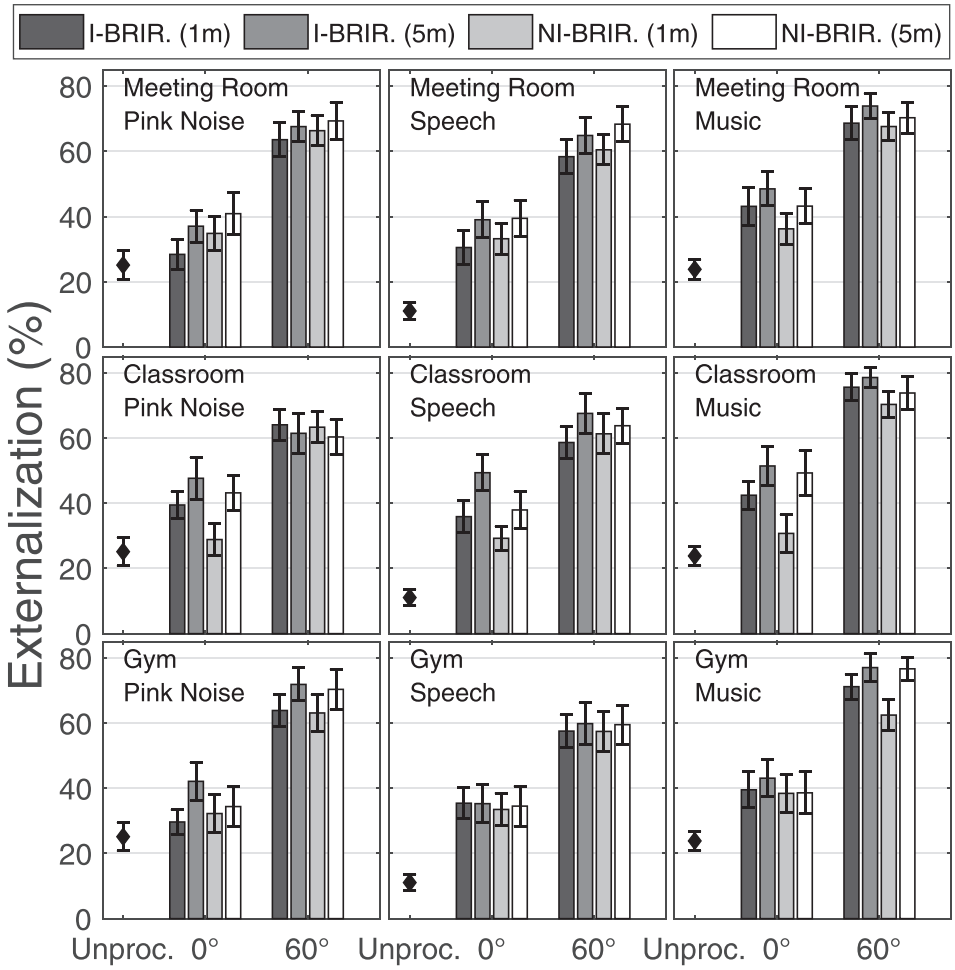

FIG. 2. Averaged degrees of perceived externalization across listeners with standard errors measured in Experiment 1 for three rooms (rows) and three signal types (columns). Within each panel, externalization scores (Y-axis) are plotted as a function of source azimuth (X-axis). The source distance and the use of individualized vs non-individualized BRIRs (I-BRIRs and NI-BRIRs, respectively) are color-coded according to the legend. At the very left of each panel, the mean externalization score with standard error obtained in the unprocessed condition is plotted (diamond symbol) and duplicated over rows (rooms) for easier comparisons.

those of the frontal source for all signals and rooms ($p < 0.002$). The influence of the room was only significant in the case of a lateral source of pink noise, which was perceived significantly more externalized in the meeting room than in the classroom ($p = 0.043$). The factor signal type was not significant at any level of the other factors (room and azimuth).

C. Discussion

In Experiment 1, the average externalization rating obtained with NI-BRIRs (51.2%) was very close to that obtained with I-BRIRs (53.4%), even if the difference was significant in the classroom and in the gym (5 and 2 percentage points). This suggests that, in the absence of a real source used as a reference, the individualization of the BRIRs only contributed marginally to perceived externalization in most of the conditions tested here. Most of the externalization cues were already present in the NI-BRIRs. This is in agreement with Cubick *et al.* (2015) who also observed only slightly higher externalization ratings with I-BRIRs compared to NI-BRIRs. Begault *et al.* (2001) did not observe any significant effect of individualization when measuring externalization in anechoic and reverberant conditions. Kim and Choi (2005) noticed a trend for improvement regarding externalization when using I-HRTFs compared to NI-HRTFs, especially for frontal sources. However, the authors did not present statistical analyses of their results and only a small number of subjects was involved (five, including the two authors). The present study did not test sources located behind the listener, which might have highlighted more pronounced differences between I-BRIRs and NI-BRIRs, as it is the case for localization (Møller *et al.*, 1996; Wenzel *et al.*, 1993).

In agreement with previous studies (Kates *et al.*, 2018; Hendrickx *et al.*, 2017; Cubick *et al.*, 2015; Hiipakka *et al.*, 2012; Kim and Choi, 2005), a strong significant effect of source azimuth was observed here. In addition to the noticeable increase of externalization due to source azimuth in the raw data (Fig. 2), simple main effect analyses always highlighted that for all levels of any other factor, a lateral source (60°) was perceived significantly more externalized compared to a frontal source (0°). While Li *et al.* (2019) did not report a significant effect of azimuth, they observed a similar trend in which externalization tended to be lower for a frontal source compared to lateral sources. However, other studies did not observe such an azimuth effect and reported similar externalized percepts for both frontal and lateral sources (Hassager *et al.*, 2016; Catic *et al.*, 2015). One major difference in those studies was the presence of a visual reference during the listening tests. Subjects had to indicate how far from this reference they perceived the sources. This type of question and the presence of the visual cue could have had two opposite effects explaining the absence of difference in terms of externalization between lateral and frontal sources (as well as the reduced non-significant trend in the study of Li *et al.*, 2019). First, the visual cue could have help subjects externalize frontal sources, as also observed in the

ventriloquism effect where sound localization is highly influenced by a visual reference (Paquier *et al.*, 2016; Hendrickx *et al.*, 2015; Radeau and Bertelson, 1977). Second, the distance question and the visual reference to match could have resulted in lateral sources being judged less externalized (than it would have been the case without the real source present) because they were potentially not perceived as far as the reference source. Another explanation could be that the azimuths tested (0° and 30° in Catic *et al.*, 2015; and 0° and 50° in Hassager *et al.*, 2016) were not large enough to observe an azimuth effect. This is partly supported by the results of Cubick *et al.* (2015) who did not observe a significant increase in externalization when varying the azimuth of the source from 0° to 25°. Alternatively, it could also be argued that listeners' ratings could have been based on lateralization and not externalization. More azimuths needed to be tested to rule out this hypothesis. This was done and is discussed in Experiment 3.

When increasing the source distance, externalization ratings increased by about 6 percentage points on average. This increase was not observed in all conditions, as indicated by the three-way interaction between distance, azimuth, and room. First, externalization significantly increased with distance in the presence of a frontal source in the classroom and meeting room. This supports the idea that reverberation increases externalization (Catic *et al.*, 2015; Begault *et al.*, 2001). In the presence of a frontal source located at 1-m in a room, the signals at the listener's ears are mainly composed of the direct sound and are thus highly interaurally correlated. When increasing the source distance, the contribution of the direct sound is reduced at the ears and the contribution of reverberation increases. Dichotic reflections reaching the ears provide helpful dynamic binaural cues regarding externalization (Catic *et al.*, 2013; Catic *et al.*, 2015). When considering lateral sources, the 60° azimuth also caused a decorrelation between left and right signals, resulting in higher externalization for all rooms in comparison to the frontal sources. This decorrelation could have been large enough that increasing the distance did not result in further increases of externalization for the classroom and meeting room. In the gym, a greater distance slightly but significantly increased the externalization of the lateral source. The increase of externalization with distance was not significant for the frontal source in the gym.

The statistical analyses revealed a small but significant effect of the type of signal produced by the source: the music was on average perceived more externalized than speech (by 9 percentage points) or pink noise (by 6 percentage points). However, the simple effect analysis on the azimuth × room × signal interaction did not reveal any significant effect of the type of signal at any level of azimuth or room, which suggests that the main effect of signal could also be due to residual variance. A separate one-way ANOVA with repeated measures was conducted on the unprocessed conditions. The analysis showed a main effect of the type of signal [$F(2,34) = 11.67$; $p < 0.001$]. Further Post-Hoc test using the Bonferroni corrections indicated that the speech signal was perceived less externalized than pink noise and music. We cannot explain these small differences across signal types, which seem to vary depending on the conditions considered.

IV. EXPERIMENT 2: INFLUENCE OF THE TRANSDUCER ON EXTERNALIZATION

A. Aim and design

When there is no real source to match in a virtual simulation with headphones, inverting the frequency response of these headphones might not be so crucial: the headphones alter the spectrum of the stimuli, but it is as if the simulated source had a slightly different spectrum to start with. In the absence of a real spectrum to match, these slight differences might not be so important regarding the perceived degree of externalization produced by the simulation. This hypothesis on the influence of headphone equalization was tested in Experiment 2 using again the NI-BRIRs measured in Experiment 1.

Compared to headphones, earphones are inserted at the entrance of the ear canal, so that there is no filtering of the sound by the pinnae. Moreover, wearing earphones "feels" different for a listener, as he/she does not have the weight of the headphones on their head. For these reasons, Experiment 2 also compared the perceived degrees of externalization measured with earphones and headphones (without headphone equalization).

Three conditions regarding the transducers were tested in Experiment 2 throughout three experimental blocks: headphones with equalization (as in Experiment 1), headphones without equalization, and earphones (without equalization). Within each block, one pair of transducers was tested while randomly varying the azimuth of the source (0° and 60°), its distance (1 m and 5 m), and the type of signal reproduced (pink noise, speech, music). Within each block, each type of signal was also presented diotically without any filtering (unprocessed conditions). Each block then resulted in 15 stimuli randomly presented with four repetitions. Across the 18 listeners, all six possible permutations of the three blocks were tested to counterbalance any order effect related to the transducers (each block permutation was then tested by three listeners). Only the NI-BRIRs measured in the meeting room were used in Experiment 2.

The practice session of Experiment 2 was identical to the one of Experiment 1.

B. Results

Figure 3 presents the mean degrees of perceived externalization measured in Experiment 2. For every transducer, unprocessed stimuli resulted in lower externalization compared to convolved stimuli. The results indicate that neither inverting the frequency response of the headphones nor switching from headphones to earphones had an effect on externalization scores.

A repeated-measures ANOVA with four within-subjects factors (azimuth × distance × signal type × transducer) was performed on the data of Experiment 2 (omitting the unprocessed conditions). There were significant main effects of azimuth [$F(1,17) = 94.45$; $p < 0.001$] and distance [$F(1,17) = 10.39$; $p = 0.005$]. Two-way interactions were also significant: transducer × distance [$F(2,34) = 4.62$; $p = 0.017$] and azimuth × distance [$F(1,17) = 15.45$; $p = 0.001$]. One three-

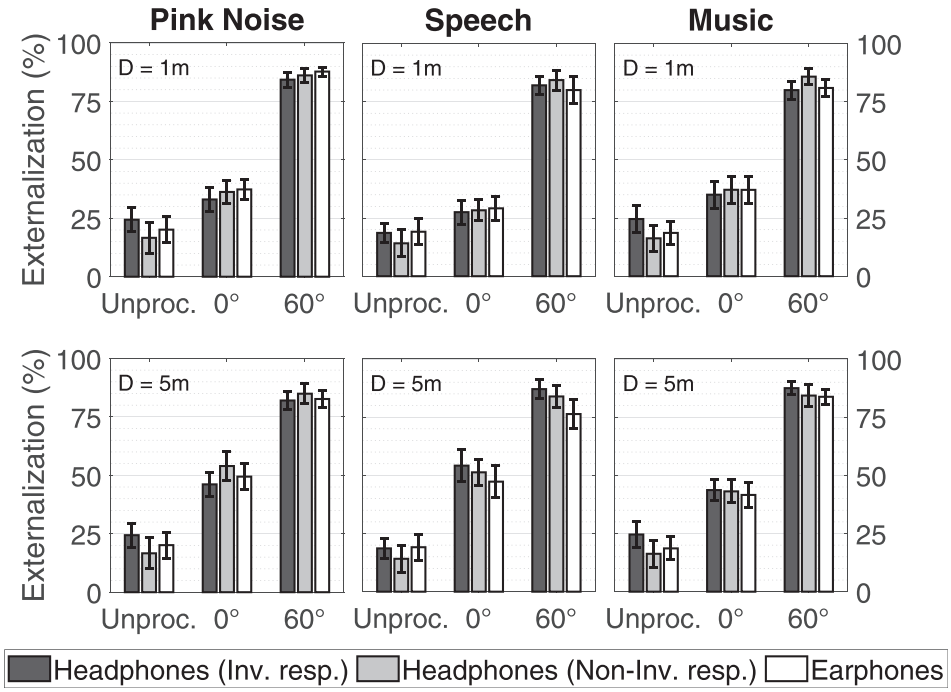

FIG. 3. Averaged degrees of perceived externalization across listeners with standard errors measured in Experiment 2 for three types of signal (columns) and two source distances (D; rows) using the NI-BRIRs from the meeting room. Within each panel, externalization scores (Y-axis) are plotted as a function of source azimuth (X-axis). Each color of the bar graph corresponds to a given transducer condition (see legend). The unprocessed condition is presented at the very left of each panel and duplicated over rows (distance) for easier comparisons.

way interaction was found significant: azimuth × distance × signal type [$F(2,34) = 5.98$; $p = 0.006$].

These significant interactions were further analyzed through simple effect analyses with Bonferroni corrections. Concerning the transducer × distance interaction, no effect of the transducers was found significant at any distance. However, sources simulated at 5 m were perceived more externalized compared to 1-m sources when heard over headphones (with and without headphone equalization; $p < 0.05$). Distance interacted with azimuth, such that an increase of distance led to an increase of externalization only when the azimuth was 0° ($p = 0.002$). The effect of azimuth was significant at all distances. The three-way interaction between azimuth, distance, and signal type provides more insights on the previously mentioned two-way interaction by showing that distance had a significant effect on externalization scores only for a frontal pink noise source ($p = 0.004$) or a frontal speech source ($p = 0.003$). Again, the effect of azimuth was significant at all distances and for all signal types ($p < 0.001$), whereas the signal type was never found significant at any azimuth or distance.

C. Discussion

As in Experiment 1, a strong effect of azimuth was observed in Experiment 2. Increasing the source azimuth from 0° to 60° systematically resulted in an increase of externalization for all levels of the other factors. Statistical analyses found a significant interaction between azimuth, distance, and signal type. The effect of azimuth was significant at all distances and for all signal types. On the other hand, the type of signal had no significant effect at any distance or azimuth. This three-way interaction can then be summarized by the significant effect of distance observed only for the pink noise and speech signals for frontal sources. This effect was already observed in Experiment 1 for the same room (meeting room), except that it did not depend on the type of signal.

Some conditions from Experiment 2 can be directly compared to some conditions of Experiment 1 because identical stimuli have been used. Namely, results in the equalized conditions at 0° in Experiment 2 were compared to the corresponding conditions in Experiment 1 (i.e., meeting room, NI-BRIR, 0°). Externalization was higher in Experiment 1 for distances of 1 m, and higher in Experiment 2 for distances of 5 m. However, the difference across experiments was not significant [$|t(34)| < 1.67$; $p > 0.10$]. Results from Experiment 1 were then consistently reproduced in Experiment 2 with different listeners.

The conditions tested in Experiment 2 found no significant effect of the transducers: neither inverting the frequency response of the headphones nor switching from headphones to earphones significantly affected perceived externalization. This was also observed by Hiipakka *et al.* (2012), who compared binaural renderings in terms of localization and externalization in the absence of visual cues and across different methods of reproduction including headphones vs earphones, and equalized vs non-equalized. They measured externalization with a binary scale (the subject ticked whether the sound was perceived inside or outside the head), and no significant difference was observed between the reproduction methods. While here the transducer significantly interacted with the distance factor, the simple main effect analysis indicated that the transducer did not significantly affect externalization at any source distance. It should be repeated here that, when the aim of the simulation over headphones is to match a real source, the spectrum alterations caused by the frequency response of the headphones could potentially change the perceived timbre of the simulated source compared to the one of the real source. However, when there is no real source spectrum to match, headphone equalization is not an absolute requirement to provide an externalized percept to the listener and the choice of the transducers might not be so crucial to achieve a reasonable externalization.

The results of Experiment 2 support the robustness of externalization to spectral modifications previously observed by Kulkarni and Colburn (1998) and Hassager *et al*. (2016). Kulkarni and Colburn used a discrimination paradigm in anechoic conditions in which listeners had to indicate if the sound they heard came from a visible loudspeaker or from a tube-phone apparatus they wore at all time. They examined the robustness of individualized HRTFs on real/virtual confusion by testing different levels of HRTFs deterioration. Their results indicated that listeners only distinguished virtual from real sources at extreme levels of deterioration, which caused a difference in perceived elevation and was used as a discrimination cue. Hassager *et al*. confirmed the results of Kulkarni and Colburn while controlling the amount of spectral detail in the direct part of individualized BRIRs (mostly influenced by the HRTFs). They also showed that externalization remained very high when the spectral detail was reduced only in the reverberant part of the BRIR.

The present study highlighted that externalization perceived through earphones did not significantly differ from that perceived through headphones, at least for the conditions tested. One could have expected different externalization ratings with earphones compared to headphones due to the direct placement of the transducer into the ear canal, then by-passing the acoustical coupling between the headphone and the ear canal causing resonances in the pinna cavities (Kulkarni and Colburn, 2000). The present results suggest that such a coupling did not influence externalization in our study, or that earphones induce a different coupling.

V. EXPERIMENT 3: RELATIVE CONTRIBUTIONS OF REVERBERATION AND INTERAURAL DIFFERENCES ON EXTERNALIZATION

A. Aim and design

Experiment 3 further explored the acoustical cues present in non-individualized stimuli at the listeners ears that are responsible for the perceived externalization of a virtual sound source (in the absence of a real source to match). It was aimed at better determining the role played by interaural differences and reverberation.

Perceived externalization was evaluated using four types of signal (pink noise, music, speech, clinging bottles) convolved with NI-BRIRs measured at four azimuths (0°, 30°, 60°, and 90°) in five rooms (including one anechoic). The anechoic room was considered here to highlight the influence of reverberation on perceived externalization. To better understand the nature of the cues provided by reverberation, diotic-convolved stimuli were also created by convolving the signal with the temporal average of the left/right signals of the original BRIRs, resulting in diotic signals containing spectral and temporal modifications caused by the room and the head but without any interaural difference.[5] These diotic-convolved stimuli were considered only for two rooms (Surrey C and Salford 2 m) and two azimuths (30° and 90°). As control conditions, unprocessed diotic signals (not convolved) were also tested, resulting in a total of 100 conditions (4 unprocessed, 16 diotic-convolved, and 80 convolved), which were measured twice for each listener.

The listeners first took part into a practice session of 12 trials. The four types of signal were randomly presented unprocessed, convolved with Salford2 (at 0°) and convolved with Surrey C (at 90°).

B. Results

Figure 4 presents the mean degrees of perceived externalization measured in the convolved conditions of Experiment 3. Again, listeners reported a much stronger externalization when listening to a sound convolved with NI-BRIRs compared to unprocessed signals. Externalization scores increased for lateral sources compared to frontal sources. Scores were lower when sounds were convolved with the anechoic BRIRs compared to any reverberant BRIR, especially for lateral source positions.

A repeated-measures ANOVA was performed on the convolved conditions with three within-subjects factors (signal type, azimuth, room). Significant effects were found for the room [$F(4,80) = 18.12$; $p < 0.05$] and for the azimuth [$F(3,80) = 17.71$; $p < 0.05$]. *Post hoc* tests (HSD Tukey) on these main effects indicated that the anechoic room produced significantly lower externalization than all the other rooms, and that the scores obtained at 0° were significantly lower than those obtained at any other azimuth. No interaction was found to be significant.

Figure 5 compares the externalization ratings obtained in four listening conditions: unprocessed, diotic-convolved, convolved with anechoic BRIRs and convolved with reverberant BRIRs. The data is averaged across subjects and signal types. Since diotic-convolved conditions were only tested with two azimuths (30° and 90°) and two rooms (Surrey C and Salford

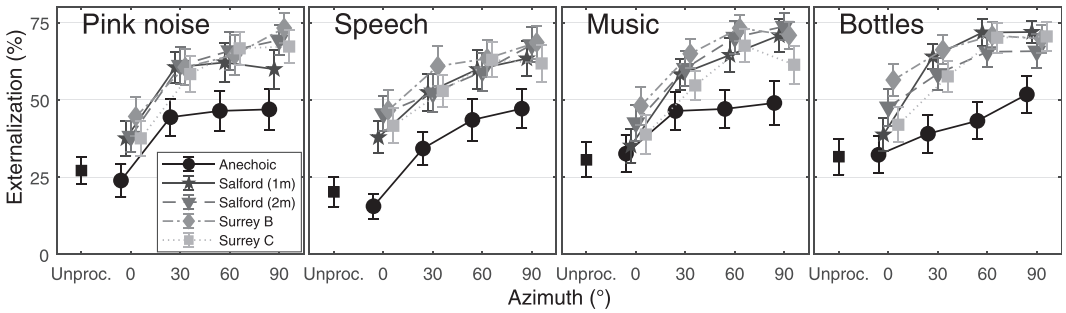

FIG. 4. Averaged degrees of perceived externalization across listeners with standard errors for the convolved conditions measured in Experiment 3 as a function of azimuth (X-axis), signal type (panels) and room (lines). Averaged values obtained in the unprocessed conditions (anechoic, diotic) are plotted with standard errors at the left of each panel (black squares).

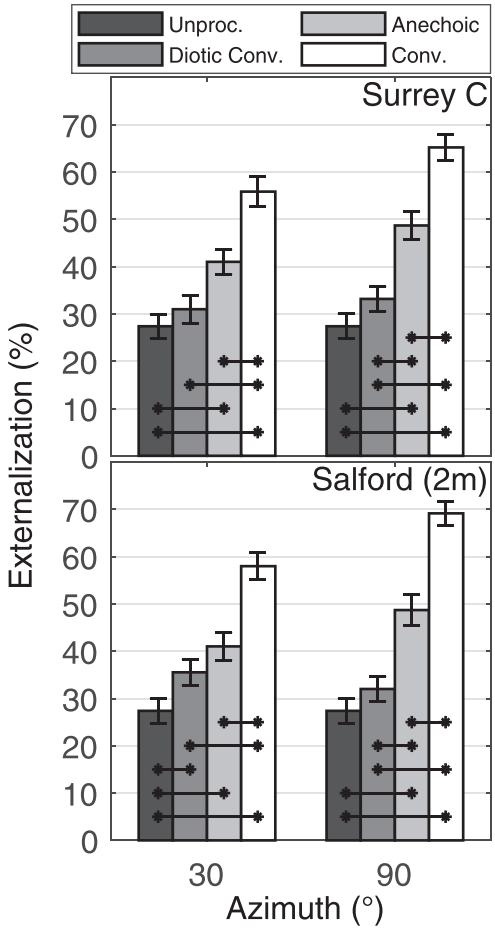

FIG. 5. Averaged degrees of perceived externalization obtained with the stimuli which were unprocessed, diotic-convolved (Diotic Conv.), convolved with anechoic BRIRs (Anechoic), or convolved with reverberant BRIRs (Conv.), as measured in Experiment 3 for two rooms (panels) and two azimuths (X-axis). Scores are averaged across listeners and types of signal; vertical lines represent standard errors. Horizontal lines connect pairs of significantly different scores ($p < 0.05$). The score from the unprocessed condition was duplicated in each cluster of bars, and scores from the anechoic conditions were duplicated across room panels.

2), the data from the convolved conditions (dichotic BRIRs) is also limited to these conditions in Fig. 5. In the diotic-convolved conditions, externalization was strongly reduced compared to the dichotic-convolved conditions. Anechoic interaural differences increased externalization significantly compared to the unprocessed conditions, while diotic reverberation almost never demonstrated a significant influence. In combination, however, interaural differences and reverberation caused a greater score increase (between 28 and 40 percentage points).

For each room and azimuth, listening modes were compared through paired-samples t-tests with Bonferroni corrections. Connection lines in Fig. 5 indicate significant differences between the corresponding listening conditions. Convolved conditions (white bars, Fig. 5) resulted in a significantly higher externalization score than all other conditions [$t(83) > 4.21$; $p < 0.01$]. When azimuth was 90°, using anechoic BRIRs led to better externalization compared to diotic signals (either unprocessed [$t(83) > 4.9$; $p < 0.01$], or convolved with the same impulse response for both ears [$t(83) > 3.4$; $p < 0.01$]). Finally, monaural reverberation (diotic-convolved) did not significantly improve externalization compared to unprocessed signals {except in Salford 2 at 30° [$t(83) = 3.58$; $p < 0.01$]}.

C. Discussion

Experiment 3 first confirmed the results of Experiments 1 and 2. Perceived externalization was generally improved when listening to a sound convolved with non-individualized BRIRs compared to unprocessed signals (except in the case of an anechoic frontal source, for which the ear signals are as diotic as the unprocessed signals). Lateral sources were perceived as more externalized compared to frontal sources, even in the anechoic room. On average, externalization ratings significantly increased when increasing the source azimuth from 0° to 30°, but larger azimuths did not result in significantly higher ratings: not being in front of the listener improved the externalization of the source, no matter how lateral the source was, and even in the absence of reverberation. The plateau observed from 30° confirms that listeners did not rate lateralization instead of externalization, because if they had done so, it would have been expected that ratings keep increasing with azimuth. These results confirm the findings reported by Kates *et al.* (2018) who observed a similar influence of the azimuth on externalization for the same angles (increase and plateau) in the presence of visual references. They also tested a "pan-pot" condition where only differences of level were used between the ear to lateralize sounds. Their "pan-pot" condition at 0° is therefore directly comparable to our present diotic "unprocessed" conditions. Similar to the results presented here, using I-HRIR or I-BRIR (with simulated reverberation) always resulted in higher externalization compared to "pan-pot" on average across subjects and azimuths.

The sources simulated in the anechoic room were less externalized than those simulated in the reverberant rooms, especially for lateral sources, which confirmed the importance of room reflections for externalization (Kates *et al.*, 2018; Catic *et al.*, 2015; Begault *et al.*, 2001). Catic *et al.* (2015) varied the energy of reflections present in their BRIR by truncating the BRIR at different early/late limit (from 2.5 up to 500 ms). Perceived externalization increased when increasing the early/late limit, indicating that listeners relied on cues contained in the late part of the BRIR to externalize. Interestingly, when Hassager *et al.* (2016) reduced the spectral detail in the late part of the BRIR, externalization was not affected. These two results do not necessarily conflict with each other but rather highlight that externalization is (at least partially) related to the temporal (rather than spectral) aspect of late reflections of BRIRs.

Experiment 3 also showed that perceived externalization was strongly reduced when the interaural differences created by reverberation were eliminated (diotic-convolved conditions, Fig. 5) compared to when they were present (convolved conditions). Thus, reverberation improved externalization only if it produced interaural differences. This key aspect is in agreement with Catic *et al.* (2015) who also observed that diotic late reflections resulted in lower externalization compared to binaural late reflections. By testing anechoic conditions, the present study also highlighted the synergetic relationship between interaural differences and reverberation. In

isolation, anechoic interaural differences improved externalization compared to the unprocessed condition (Fig. 5), while diotic reverberation did not have a strong influence. In combination, however, these two factors caused a greater increase of externalization ratings, suggesting that externalization relied on a binaural cue associated with reverberation.

Externalization was not influenced by the type of signal produced by the source in Experiment 3.

VI. ACOUSTICAL CORRELATES OF EXTERNALIZATION WITHOUT REFERENCE

A. Aim and design

Because Experiment 3 provided evidence that perceived externalization might rely on reverberation-related interaural characteristics (Fig. 5), all the stimuli from all experiments were analyzed according to the method described below (Sec. VI B). The interaural coherence (IC), interaural time difference (ITD), and ILD of the stimuli were computed using different temporal and spectral resolutions (long- or short-term, broadband, or within narrow frequency bands). The average and standard deviation of the resulting three attribute distributions were used to compute their correlation with the externalization scores across the tested conditions of each experiment.

B. Signal analyses

In order to identify acoustical correlates of the externalization ratings, all the stimuli were analyzed using different temporal and spectral resolutions: long- or short-term, and broadband or within narrow frequency bands. For each stimulus, and for each frequency/time scale, the interaural coherence, the ITD and the ILD were computed. When the analysis was performed in multiple frequency bands, or in multiple time frames (or both), it resulted in a distribution of binaural indicators. Both the average and standard deviation of these distributions were used to evaluate their correlation with the externalization scores.

A normalized interaural cross-correlation function (ICF) was first computed using the left and right channels of each stimulus, where the lag range was limited to [−1, 1] ms. The interaural coherence was determined as the maximum value of the ICF, while the ITD was defined as the lag value at which this maximum occurred. The ILD was computed separately by taking the energy ratio between the left and right ear signals.

Narrow-band analyses were conducted by using a fourth-order gammatone filterbank developed by Hohmann (2002) and implemented in the Auditory Model Toolbox (AMT; Søndergaard and Majdak, 2013). The filterbank was composed of 33 bandpass filters with center frequencies linearly spaced on the equivalent rectangular bandwidth scale (ERB; Glasberg and Moore, 1990) from 73 Hz up to 9.3 kHz with a bandwidth of 1 ERB. For short-term analyses, the signals were segmented into 50-ms time frames with a 5-ms overlap using a Hann window.

C. Results

Figure 6 presents the Pearson correlation coefficients obtained between the absolute value of each indicator and the externalization scores of Experiments 1 to 3 (rows). Star symbols indicate significance of the correlation accounting for Bonferroni corrections. These results systematically indicated strong and significant correlations between externalization score and most binaural cues. The broadband long-term ITD was highly correlated with the externalization score in the three experiments (between 0.79 and 0.95). However, when looking at the linear regressions (not shown here), the data points were not spread along the regression lines but rather grouped as clusters due to the discrete azimuths tested. These high correlations are due to the strong effect of azimuth on externalization, but the ITD cannot explain the variations of externalization due to other factors, such as

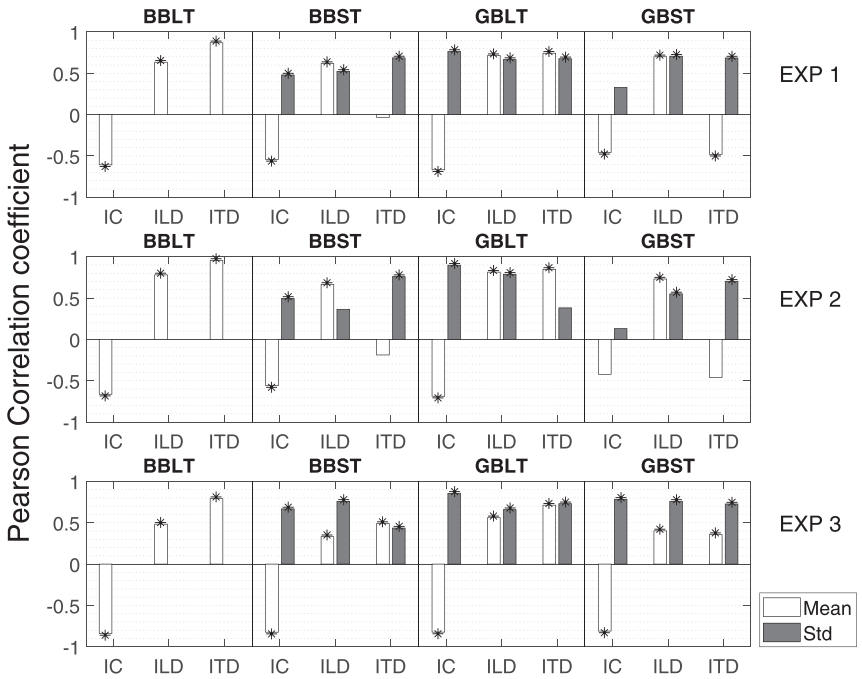

FIG. 6. Pearson correlation coefficients obtained between three binaural cues (interaural coherence IC, ILD, ITD; X-axis) and the mean degree of perceived externalization measured in each experiment (rows), when considering four spectro-temporal resolutions (columns): Broadband/Gammatone bands (BB/GB), and Long-term/Short-term (LT/ST). When the analysis involved gammatone bands or short-term resolution (or both), either the mean (in white), or the standard deviation (in gray) of the cues was considered to compute the correlation.

reverberation. The standard deviation of the narrowband long-term IC can predict those variations and showed a systematically high correlation with externalization across experiments (0.76 in Experiment 1, 0.89 in Experiment 2, and 0.85 in Experiment 3). Overall, it was the best attribute to describe the externalization ratings of the present study. However, many other binaural cues showed significant correlations, and it should be kept in mind that the indices considered here are correlated with each other. Another interesting result is that the broadband long-term IC, a very simple binaural attribute, always significantly correlated with externalization scores (between 0.6 and 0.84).

D. Discussion

Catic *et al.* (2015) found that externalization could be closely related to temporal changes of the IC or ILD. By plotting the 80th percentile and standard deviation of the IC and ILD, respectively, in each condition, they observed qualitatively similar curves as the externalization ratings provided by the listeners. No statistical or quantitative indication was given as how strong and significant this connection was. As they only showed one frequency band, it is also unclear whether this close correspondence stands in only a particular band or across several bands. Li *et al.* (2018) also analyzed dynamic binaural cues (IC and ILD) in a similar way as Catic *et al.* (2015) and Hassager *et al.* (2016). They showed that perceived externalization could be well predicted by fluctuations of ILD and IC. In the present study, the standard deviation (instead of the percentiles) of the IC computed within short time frames and narrow bands only correlated significantly with externalization in Experiment 3. But other temporal and spectral resolutions of the IC led to very high correlations across experiments, suggesting that, in agreement with Li *et al.* (2019) and Catic *et al.* (2015), IC could be a strong cue for externalization, even in the absence of visual reference. Concerning temporal fluctuations of the ILD, the closest indicator computed in the present study would be the standard deviation of the broadband short-term (BBST) ILD, which also showed a good correlation with externalization, but not as high as when considering narrow bands. It should also be noted that the computation of the ILD slightly differed between the present study and the one of Catic *et al.* (2015).[6]

Across three experiments conducted in the present study with different listeners and including seven rooms, many binaural cues were well correlated with externalization scores. This indicates that perceived externalization could be well estimated with very simple acoustical binaural attributes. While the aim of the present study was not to present an accurate externalization model, the correlation analyses confirmed and quantified strong evidence for the binaural nature of the externalization percept highlighted in previous studies (Hassager *et al.*, 2016; Catic *et al.*, 2013; Catic *et al.*, 2015; Hartmann and Wittenberg, 1996). The standard deviation of the long-term IC across narrow bands was particularly highly correlated to externalization in all the experiments conducted here (>0.75). This correlation was positive, meaning that the larger the variation of IC across frequency bands, the higher the perceived externalization. The broadband long-term IC was also well correlated with externalization in all experiments (>0.6). This could represent a very simple way to obtain a first estimate of perceived externalization, compared to more complicated approaches involving echo suppression, gammatone filtering, or inner hair cell modelling (Li *et al.*, 2019). These outcomes coherently suggested that externalization highly depended on the difference/similarity between the left and right signals delivered to the listeners' ears. When considering the unprocessed and diotic convolved conditions of Experiment 3, the left/right signals were identical, resulting in an IC value of 1 and in the lowest externalization scores of the present study. In the presence of a frontal source in anechoic conditions, interaural differences are very small and led to poor externalization, which could be increased by either increasing the azimuth of the source or adding room reflections (different at the two ears), causing a decrease of IC.

VII. CONCLUSIONS

The present study investigated the degree of perceived externalization of sound sources simulated over headphones by convolving different signals with individualized or non-individualized BRIRs measured in seven rooms for multiple source azimuths and source distances. Perceived externalization was evaluated here without a real source used as a reference, without visual cues (eyes closed), and without dynamic binaural rendering. The aim was to better determine the externalization that could be attributed to auditory stimulation only. The main results were:

(1) In the absence of a real source position/distance to match, perceived externalization obtained with non-individualized BRIRs was comparable to that obtained with individualized BRIRs.
(2) In the absence of a real source spectrum to match, headphone equalization did not improve externalization. The same degree of externalization was also achieved through non-equalized earphones.
(3) Lateral sources were more externalized than frontal sources.
(4) Reverberation improved source externalization, but only when it created interaural differences.
(5) Externalization ratings were strongly and negatively correlated with interaural coherence, indicating that the lower the coherence, the greater the externalization. Whether the coherence should be computed in the direct or reverberant parts (or both taken together) of the signals remains to be investigated.
(6) The type of signal produced by the virtual source did not strongly influence its perceived externalization.

ACKNOWLEDGMENTS

The authors would like to thank Virginia Best (Boston University) for suggestions on a previous version of the manuscript, Barry Clinch (National Acoustic Laboratories, Sydney) for advices on recording equipment, Jacques Grange and John Culling (Cardiff University) for advice on

the BRIR recordings, Kevin Perreaut for preparing the stimuli and collecting the data of Experiment 3, Joachim Blanc-Gonnet for advice on programming the experimental interface, and all listeners who took part in the experiments. This work was performed within the LabEx CeLyA (ANR-10-LABX-0060/ANR-16-IDEX-0005) and supported by the grant CogniComa (ANR-14-CE-15-0013).

APPENDIX

The instructions given to each listener at the beginning of an experimental session were: "Close your eyes and keep your head still. Hit the space bar to play the sound. You will hear either speech, music, noise or even bottles which will sound more or less outside your head: sometimes it will sound inside your head, and sometimes it will sound outside, as if it was a natural source around you, and sometimes it will sound "in between." You can play the same sound as many times as you want but the more spontaneous, the better. Then you open your eyes, and we ask you to indicate with a cursor your sensation of externalization between completely externalized and internalized. So focus on the "in/out of the head" aspect regardless of where and how it sounds."

[1]The linear phase filters preserved the phase spectrum despite some potential occurrences of pre-ringing artefacts, which we assumed would not affect externalization.

[2]See http://iosr.uk/software/#BRIRs

[3]See http://usir.salford.ac.uk/30868/

[4]During the experiment, there was no supervision that the listeners were closing their eyes. However, they reported doing so when informally asked after the experiment.

[5]While this type of process could have led to two peaks in the early part of the averaged BRIR (corresponding to the direct sound of each side), this was not observed here. Because of the ILD, the contralateral signal had a smaller amplitude compared to the ipsilateral signal, resulting in a high correlation between the averaged signal and the ipsilateral signal. The consequence seemed more apparent in the frequency domain with the observation of some ripple at low frequencies (below 500 Hz) for some of the processed BRIRs. We were not aware of a perfect method to design diotic stimuli with the entire information from the BRIR; however, we do not think this slight spectral artefact had a major influence on the externalization ratings collected, which were primarily correlated with binaural attributes as revealed in Sec. VI

[6]Catic *et al.* (2015) computed the ILD at the time where the IC is maximum with more processing stages (such as echo suppression and half-wave rectification), whereas the present study considered the energy of the signal within a given time frame (short or long).

Begault, D. R., Wenzel, E. M., and 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," J. Audio Eng. Soc. **49**, 904–916.

Brimijoin, W. O., Boyd, A. W., and Akeroyd, M. A. (**2013**). "The contribution of head movement to the externalization and internalization of sounds," PLoS One **8**, e83068.

Calcagno, E. R., Abregú, E. L., Eguía, M. C., and Vergara, R. (**2012**). "The role of vision in auditory distance perception," Perception **41**, 175–192.

Catic, J., Santurette, S., Buchholz, J., Gran, F., and Dau, T. (**2013**). "The effect of interaural-level-difference fluctuations on the externalization of sound," J. Acoust. Soc. Am. **134**, 1232–1241.

Catic, J., Santurette, S., and Dau, T. (**2015**). "The role of reverberation-related binaural cues in the externalization of speech," J. Acoust. Soc. Am. **138**, 1154–1167.

Cubick, J., Rodríguez, C. S., Song, W., and MacDonald, E. N. (**2015**). "Comparison of binaural microphones for externalization of sounds," in *Proceedings of International Conference on Spatial Audio 2015*, July 8–10, Graz, Austria.

Farina, A. (**2007**). "Advancements in impulse response measurements by sine sweeps," in *Proceedings of the 122nd AES Convention*, May 5–8, Vienna, Austria.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

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

Hassager, H. G., Gran, F., and Dau, T. (**2016**). "The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment," J. Acoust. Soc. Am. **139**, 2992–3000.

Hendrickx, E., Paquier, M., Koehl, V., and Palacino, J. (**2015**). "Ventriloquism effect with sound stimuli varying in both azimuth and elevation," J. Acoust. Soc. Am. **138**, 3686–3697.

Hendrickx, E., Stitt, P., Messonnier, J.-C., Lyzwa, J.-M., Katz, B. F., and de Boishéraud, C. (**2017**). "Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis," J. Acoust. Soc. Am. **141**, 2011–2023.

Hiipakka, M., Takanen, M., Symeon, D.-M., Archontis, P., and Pulkki, V. (**2012**). "Localization in Binaural Reproduction with Insert Headphones," in Proceedings of the 132nd Audio Engineering Society Convention, April 26–29, Budapest, Hungary, pp. 838–846.

Hohmann, V. (**2002**). "Frequency analysis and synthesis using a Gammatone filterbank," Acust. Acta Acust. **88**, 433–442.

Hummersone, C., Mason, R., and Brookes, T. (**2010**). "Dynamic precedence effect modeling for source separation in reverberant environments," IEEE Trans. Audio. Speech. Lang. Process. **18**, 1867–1871.

Kates, J. M., Arehart, K. H., Muralimanohar, R. K., and Sommerfeldt, K. (**2018**). "Externalization of remote microphone signals using a structural binaural model of the head and pinna," J. Acoust. Soc. Am. **143**, 2666–2677.

Kim, S.-M., and Choi, W. (**2005**). "On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach," J. Acoust. Soc. Am. **117**, 3657–3665.

Kulkarni, A., and Colburn, H. S. (**1998**). "Role of spectral detail in sound-source localization," Nature **396**, 747–749.

Kulkarni, A., and Colburn, H. S. (**2000**). "Variability in the characterization of the headphone transfer-function," J. Acoust. Soc. Am. **107**, 1071–1074.

Li, S., Schlieper, R., and 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," J. Acoust. Soc. Am. **144**, 966–980.

Li, S., Schlieper, R., and Peissig, J. (**2019**). "The role of reverberation and magnitude spectra of direct parts in contralateral and ipsilateral ear signals on perceived externalization," Appl. Sci. **9**, 460.

Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (**1996**). "Binaural technique: Do we need individual recordings?," J. Audio Eng. Soc. **44**, 451–469.

Oppenheim, A. V., and Schafer, R. W. (**2014**). *Discrete-Time Signal Processing* (Pearson, Noida, India).

Paquier, M., Côté, N., Devillers, F., and Koehl, V. (**2016**). "Interaction between auditory and visual perceptions on distance estimations in a virtual environment," Appl. Acoust. **105**, 186–199.

Radeau, M., and Bertelson, P. (**1977**). "Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations," Percept. Psychophys. **22**, 137–146.

Satongar, D., Lam, Y. W., and Pike, C. (**2014**). "Measurement and Analysis of a Spatially Sampled Binaural Room Impulse Response Dataset," in *21st International Congress on Sound and Vibration*, pp. 1–8.

Søndergaard, P., and Majdak, P. (**2013**). "The auditory modeling toolbox," in *The Technology of Binaural Listening*, edited by J. Blauert (Springer, Berlin), pp. 33–56.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123.

Wightman, F. L., and Kistler, D. J. (**1989a**). "Headphone simulation of free-field listening I: Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F. L., and Kistler, D. J. (**1989b**). "Headphone simulation of free-field listening II: Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.

Zahorik, P. (**2001**). "Estimating sound source distance with and without vision," Optom. Vis. Sci. **78**, 270–275.