Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis

Etienne Hendrickx,[1, a)] Peter Stitt,[2] Jean-Christophe Messonnier,[1] Jean-Marc Lyzwa,[1] Brian FG Katz,[3] and Catherine de Boishéraud[1]

[1)] *Conservatoire National Supérieur de Musique et de Danse de Paris, 209, avenue Jean-Jaurès, 75019 Paris, France*

[2)] *Audio Acoustics Group, LIMSI, CNRS, Université Paris-Saclay*

[3)] *Sorbonne Universités, UPMC Univ Paris 06, CNRS, Institut d'Alembert, Paris, France*

(Dated: 7 May 2017)

Binaural reproduction aims at recreating a realistic audio scene at the ears of the listener using headphones. In the real acoustic world, sound sources tend to be *externalized* (that is perceived to be emanating from a source out in the world) rather than *internalized* (that is perceived to be emanating from inside the head). Unfortunately, several studies report a collapse of externalization, especially with frontal and rear virtual sources, when listening to binaural content using non-individualized Head-Related Transfer Functions (HRTFs). The present study examines whether or not head movements coupled with a head tracking device can compensate for this collapse. For each presentation, a speech stimulus was presented over headphones at different azimuths, using several intermixed sets of non-individualized HRTFs for the binaural rendering. The head tracker could either be active or inactive, and the subjects could either be asked to rotate their heads or to keep them as stationary as possible. After each presentation, subjects reported to what extent the stimulus had been externalized. In contrast to several previous studies, results showed that head movements can substantially enhance externalization, especially for frontal and rear sources, and that externalization can persist once the subject has stopped moving his/her head.

PACS numbers: 43.66.Pn
Keywords: Binaural hearing, Externalization, Head tracking, Non-individualized HRTFs

I. INTRODUCTION

Binaural rendering uses headphones to (re)create an audio scene at the ears of a listener, by producing as accurately as possible at the listener's eardrums the waveforms that would have been produced by real stimuli at the same positions. Individualized binaural recordings can be achieved in two different ways, either naturally or synthetically. In natural recordings, real sound sources are recorded with microphones placed in the ears of the listener. In synthetic recordings, rather than record real stimuli directly, the acoustical transfer functions, from free-field to the listener's eardrums, are measured at many source positions and incorporated as digital filters which are then used to synthesize stimuli. This set of transfer functions is termed the Head-Related Transfer Function (HRTF). It includes the primary localization cues: interaural time differences (ITDs), interaural level differences (ILDs), and the monaural spectral cues.

In some applications involving binaural reproduction, it may be critical for the localization of virtual sources in direction (azimuth and elevation) to be as accurate as with real sources. The virtual sources should also be externalized rather than internalized. In other words, virtual sources should appear to originate from a source out in the world (as in real life) rather than from somewhere inside the head (Hartmann and Wittenberg, 1996; Durlach *et al.*, 1992).

Previous studies have shown that when individualized HRTFs are accurately simulated with headphones, subjects report *externalized* sources and localization accuracy comparable with free-field stimuli (Wightman and Kistler, 1989).

A. Individualized vs. non-individualized HRTF

HRTFs are strongly determined by the filtering properties of the pinnae, head, shoulders, and torso, which are specific to each individual (Wenzel *et al.*, 1993) with HRTFs varying considerably among individuals (Begault and Wenzel, 1993). If subjects listen to a binaural stimulus that is non-individualized (*i.e.* recorded with microphones placed in the ears of another individual or manikin, or synthesized using HRTFs from another individual or manikin), they may perceive the audio scene inadequately: sound sources may be poorly externalized, diffuse, or incorrectly localized. Moreover, front-back confusions might occur frequently (Hartmann and Wittenberg, 1996). Perceptual attributes linked to HRTF variations have been recently detailed in Simon *et al.* (2016).

Unfortunately, it may not be feasible in practice to measure the HRTF of each potential user of a binaural

[a)] etienne.hendrickx@univ-brest.fr

rendering system (Wenzel *et al.*, 1993; Katz and Parseihian, 2012), as it can be a complex and expensive process (Mendonça *et al.*, 2012). It is therefore critical to determine to what extent the general population of listeners can obtain 1) adequate localization cues and 2) sufficient externalization when using non-individualized HRTFs.

Previously, Wenzel *et al.* (1993) asked 16 subjects to judge the apparent direction of wideband noise bursts presented in the free-field or over headphones. Results showed that localization of virtual sources was quite accurate and comparable to free-field sources for most subjects, even though non-individualized HRTFs were used. However, many subjects exhibited higher rates of front-back and up-down confusions with virtual sources compared to free-field stimuli. For speech stimuli reproduced in the horizontal plane, Begault *et al.* (2001) observed that individualized HRTFs offered no advantage in localization accuracy.

Several studies have investigated externalization using non-individualized HRTFs. According to Hartmann and Wittenberg (1996), the synthesis of a distant source leads to a perfectly externalized image if the HRTFs are properly individualized, whereas it leads to an image that is often perceived on the surface of the skull if the HRTFs are non-individualized. With five subjects and short bursts of white noise reproduced in the horizontal plane, Kim and Choi (2005) observed that sound sources synthesized with individualized HRTFs were perceived at a greater and more consistent distance than those synthesized with non-individualized HRTFs. On the other hand, with speech stimuli, neither Møller *et al.* (1996) nor Begault *et al.* (2001) reported a significant difference in externalization between individualized and non-individualized binaural synthesis.

B. Frontal and rear sources vs. lateral sources

Using virtual sources synthesized in the horizontal plane with non-individualized HRTFs and no head tracking, Laws and Platte (1975), Kim and Choi (2005), and Begault and Wenzel (1993) observed that lateral stimuli were almost always judged to be external, whereas frontal or rear stimuli were much more likely to be perceived inside the head. Note that Begault and Wenzel (1993) used anechoic speech stimuli.

Because lateral sources are already well externalized without head tracking, it is in the case of frontal and rear sources that head tracking can be expected to have a more beneficial impact on externalization.

C. Head tracking

In the real world, sound sources are in constant motion with respect to the listener because the head is never perfectly still (König and Sussmann, 1955). Moreover, if the listener turns his/her head, the egocentric auditory environment rotates by the corresponding amount in the opposite direction.

However, when listening to virtual sources under normal headphone presentation, the location of a source moves with the head, and a source directly to the left of the listener remains directly to the left no matter how he or she moves. This issue can be solved by coupling the binaural rendering system with a head tracking device, thus enabling the virtual sources to move appropriately to the listener's head movements.

Previous studies have shown that head movements enable subjects to localize real sources more accurately (Perrett and Noble, 1997) and reduce the number of front-back confusions (Wightman and Kistler, 1999). Head movements have been shown to be useful in distance perception of virtual sources using Wave-Field Synthesis rendering (Rébillat *et al.*, 2012). Similarly, head movements coupled with head tracking improve localization performance of virtual sources compared to normal headphone presentation (Begault *et al.*, 2001; Wightman and Kistler, 1999; Martin *et al.*, 2001; Noble, 1987).

However, the role of head movements in the phenomenon of externalization remains unclear. Some studies claim that head movements coupled with head tracking enhance externalization (Loomis *et al.*, 1990; Kawaura *et al.*, 1991). However, these studies were either informal or lacked sufficient subjects (only three subjects in Kawaura *et al.*) and quantitative data. Other studies suggest that the effect of head movements coupled with head tracking on externalization is small (Wenzel, 1995) or even null (Begault *et al.*, 2001).

In Begault *et al.* (2001), nine naïve subjects listened to brief speech stimuli (3 s long) reproduced at different azimuth positions (0°, ±45°, ±135°, 180°) with three different levels of reverberation: anechoic, early reflections only, and full reverberation (early reflections + late diffuse reverberation response, with a mid-band reverberation time of 1.5 s). Two different conditions were evaluated:

- The head tracker was active and subjects were requested to move their heads. Note that subjects were not instructed to move their heads in any particular manner (*i.e.*"freestyle" movements).
- The head tracker was inactive. It is assumed that subjects did not move their heads for that condition, as they were not requested to.

After each presentation, subjects had to provide estimates of distance via computer mouse, using an interactive graphic showing a head in top view. Results were then converted into externalization rate, defined as the percentage of time a stimulus was perceived outside the head. The edge of the head in the graphic was set at 4 inches and the cutoff point for treating a judgment as externalized was set to > 5 inches in order to yield a conservative estimate that eliminated judgments very close to the edge of the head.

Results showed that head tracking did not increase externalization, whether individualized or non-individualized HRTFs were used for the binaural rendering. However, the study acknowledged that the short duration of the stimuli may have limited the ability of the subjects to take advantage of cues derived from head movements. The fact that results were averaged across all positions before analysis may also explain why the effect of head tracking was not significant. As lateral sources are already well externalized without head tracking (see Section I B), it is rather for frontal and rear sources that head tracking can be expected to have a substantial impact. Thus, any small improvements occurring for lateral sources may statistically mask larger improvements for frontal and rear sources.

In Wenzel (1995), six subjects listened to a 3 s broadband Gaussian noise presented from 40 different locations: eight azimuths every 45° for five different elevations (−36° to +36°), using non-individualized HRTFs. Two different conditions were evaluated: (1) neither head tracking nor head movement versus (2) with head tracking and head movements (though subjects were requested not to lean their heads far forward or to the side).

After each presentation of a stimulus, subjects had to provide numerical estimates of distance in inches (the distance scale had anchors at 0 inches for a sound at the center of the head and 4 inches for a sound located at the perimeter of the head). Results were converted to externalization rate, defined as the percentage of time an estimation was > 4 inches. Note that the cutoff point for treating a judgment as externalized was slightly smaller than for Begault *et al.* (2001).

There was a general trend toward greater externalization when subjects moved their heads. However, the improvement in externalization rate was moderate (from 74.5% to 83.5%), possibly because stimuli were quite brief (3 s) and because results were averaged across all positions, as in Begault *et al.* (2001).

In Brimijoin *et al.* (2013), six subjects listened to short phrases (3 s long), reproduced in the horizontal plane at azimuths from −25° to +25°. Two kinds of transfer functions were measured in a room (RT30 = 0.35 s): individualized HRTFs and transfer functions measured from a simple pair of microphones on a bar. These "head-absent" transfer functions (HATFs) contained relevant reverberation cues, somewhat relevant ITD cues, but lacked spectral cues (as the filtering properties of the pinnae, head, and torso were not reproduced), which are thought to be crucial for externalization (Hartmann and Wittenberg, 1996). The individualized HRTFs and HATFs were then mixed using linear interpolation so as to create six sets of hybrid transfer functions ranging from purely head-absent (100% HATFs, 0% individualized HRTFs) to purely head-present (0% HATFs, 100% individualized HRTFs). For each presentation, subjects listened to a speech signal processed with a transfer function set randomly drawn from the six sets of hybrids. Subjects were either asked to keep their heads as stationary as possible or to rotate their heads gently back and forth between ±15°. The head tracker could be active or inactive. Thus, four tracking conditions were compared:[1]

SØ : static head orientation (no head movement), no head tracking.

ST : static head orientation, with head tracking.

MØ : head movements, no head tracking.

MT : head movements, with head tracking.

Conditions **SØ** and **MØ** correspond to "normal headphone" presentation while condition **MT** corresponds to a typical "headphone with head tracker" situation. Condition **ST** can seem paradoxical, yet studies have shown that the head is never perfectly still even when a subject is told to remain so, and can move in azimuth by up to 5° when unsupported (König and Sussmann, 1955). Thus, even micro-movements of the head might enhance externalization. In Wersényi (2009), emulation of small head-movements of 2° were shown to increase externalization rates for ≈ 20% of the subjects.

In contrast to Begault *et al.* (2001) and Wenzel (1995), subjects in the experiment of Brimijoin *et al.* (2013) were not asked to estimate distance after each presentation, but simply to report a binary choice of whether the stimulus emanated from either inside or outside the head.

Results showed that, with pure individualized HRTFs, externalization rates in conditions **SØ**, **ST**, and **MT** were high and comparable. In other words, head movements coupled with head tracking did not substantially enhance externalization compared to the conditions where the subject did not move his/her head. However, externalization collapsed dramatically when subjects moved their heads without head tracking (**MØ**).

With mixtures of individualized HRTFs and HATFs, head movements coupled with head tracking (**MT**) did provide more externalization than in the conditions without head movement (**SØ** and **ST**) as the proportion of HATFs in the mixtures were increasingly predominant over the proportion of individualized HRTFs. The increased externalization rate in condition **MT** was especially high for the mixture (20% Individualized HRTFs, 80% HATFs): ≈ +43% compared to the conditions without head movement (**SØ** and **ST**). As with pure individualized HRTFs, condition **MØ** always presented the lowest externalization rate.

With pure HATFs, externalization rates were globally very low for all conditions, even though head movements coupled with head tracking (condition **MT**) provided more externalization than in all the other tracking conditions (≤ +21%). This suggests that head movements coupled with head tracking might be more beneficial for externalization when the binaural synthesis is not individualized.

Whether individualized HRTFs, HATFs, or mixtures were used, results for conditions **SØ** and **ST** were very similar.

D. Summary and aim of the present study

Loomis *et al.* (1990) and Kawaura *et al.* (1991) have suggested that head movements enhance externalization when dynamic binaural rendering includes head tracking, however these studies lack quantitative data.

Other studies have concluded that this enhancement is weak or non-significant. However, these poor results might be due to the fact that results were averaged across all source positions, thus potentially masking significant enhancements for frontal and rear sources. Another reason could be that stimuli were very brief ($\leq 3\,\mathrm{s}$), thus giving subjects little time to take advantage of cues derived from head movements and to make large head movements. An informal test conducted by the authors of the current study suggested that large head movements ($\pm 90^\circ$ for example) were actually required to observe a substantial improvement in externalization.

In Brimijoin *et al.* (2013), the improvement brought by head movements and head tracking was more or less pronounced whether the binaural synthesis used individualized HRTFs, HATFs, or mixtures of individualized HRTFs and HATFs. Note that Brimijoin *et al.* did not conduct their experiment with *non*-individualized *head-related* transfer functions, which is a more generalizable display scenario than a synthesis using individualized HRTFs (indeed, it may not be feasible for everyone to have access to his/her own individualized HRTFs), and which, in contrast with HATFs, do contain spectral cues that are thought to be essential for externalization (Hartmann and Wittenberg, 1996).

Moreover, the protocols of all previous studies investigated whether head movements improve externalization *while* subjects are moving their heads (*immediate effects*). However, it is crucial to determine whether substantial improvements can still be observed once the subject has stopped moving his/her head (*aftereffects*). Indeed, the practical interest of head tracking would be severely reduced if it enables improved externalization only *while* subjects move their heads. This means that listeners would have to move their heads continuously to listen to binaural content with optimal externalization.

Another issue raised by Brimijoin *et al.* (2013) was that previous studies often lack detailed data concerning the extent and velocity of subjects' head movements (the fact that movements were "freestyle" in most studies probably made them difficult to summarize pertinently). Thus, experiments cannot be replicated accurately and comparisons of results with other studies are problematic.

The aim of the present study was to reproduce the experiment of Brimijoin *et al.* (2013) while addressing the issues raised above. Thus:

- The binaural synthesis was non-individualized instead of individualized to represent a more generalizable display scenario.
- The transfer functions used for the binaural synthesis were "head-related", because "head-absent" transfer functions include non-realistic ILD and ITD cues, while also lacking spectral cues that are thought to be essential for externalization.
- The full horizontal plane was investigated with results analyzed for each azimuth separately.
- The stimulus was longer than in previous studies ($8\,\mathrm{s}$ instead of $2-3\,\mathrm{s}$), thus providing subjects more time to take advantage of cues derived from head movements, enabling them to make larger head movements.
- The *aftereffects* of head movements rather than the *immediate effects* were investigated.
- Subjects' head movements were more tightly controlled than in previous experiments.

The following hypotheses are presented for the current study[2]:

H1 Large head movements cause a collapse of externalization when the head tracker is inactive, as Brimijoin *et al.* (2013) observed with individualized HRTFs.

H2 Large head movements improve externalization when the head tracker is active, especially for frontal and rear sources.

H3 Even when subjects are requested not to move their heads, they still make involuntary micromovements that improve externalization if the head tracker is active, but to a lesser extent than if they make large head movements.

II. EXPERIMENTAL SETUP

For each presentation, via headphones, subjects listened to an $8\,\mathrm{s}$ six-channel binaural stimulus, consisting of a male voice with surrounding reverberation channels. Using dynamic binaural rendering, the stimulus could be rotated around the subject and was thus presented at different orientations. Different interleaved non-individualized sets of HRTFs were used. Head tracking could either be active or inactive, subjects either had to make large head movements or keep their heads stationary. After each presentation, subjects reported to what extent the stimulus was externalized.

A. Stimulus

The stimulus consisted of an $8\,\mathrm{s}$ extract from the French poem "L'Albatros" by Charles Baudelaire, read by a male talker ($f_0 = 107\,\mathrm{Hz}$).

The stimulus was recorded with a six-channel equal-segment microphone array, described in Williams (1991).

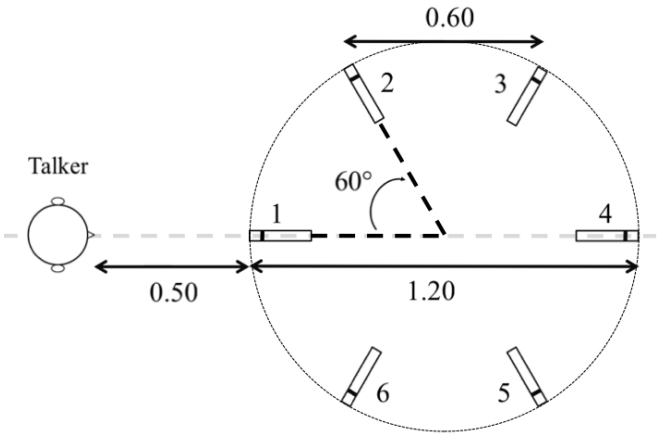

FIG. 1. Microphone array configuration used for the recording of the stimulus. Dimensions in meters.

As shown in Fig. 1, the array consisted of one front microphone (microphone 1 in Fig. 1), capturing the highest level of direct sound, and five other microphones (microphones 2–6), capturing varying levels of direct-to-reverberant energy. The microphones (cardioid directivity, DPA 4021) were arranged in a circle, 60 cm radius. The array height was 1.65 m (height of the mouth of the talker) at a distance of 50 cm to microphone 1.

It was decided to use spatial recordings with a microphone array because such arrays have been a major category of recording approaches for multichannel sound reproduction (Politis *et al.*, 2015), as they are the natural extension of the principles inherited from traditional stereophonic recording techniques. It was thus a way of presenting a binaural stimulus from a realistic system, likely to be employed in the context of real-world multichannel recording. Such microphone arrays also capture the natural reverberation, and a recent comparative study of several "binauralized" recording setups by Nicol *et al.* (2016) suggests that microphone arrays were preferred over artificial spatialization of monophonic sources using reverberation simulation. Additional details of the retained microphone array are provided in the Appendix.

The recording was made in a recording studio at the Conservatory of Paris (area of $\approx$30 m^2). The reverberation times averaged across the six microphones positions are presented in Table I. Several studies indicate that a small amount of reverberation, even in the form of a few early reflections, is sufficient to produce image externalization (Begault, 1992; Durlach *et al.*, 1992). It was thus decided not to record the stimulus in a room with too much reverberation, otherwise externalization rates may have been high, whether or not head tracking was active, potentially minimizing the influence of head tracking.

Octave band (Hz)	125	250	500	1000	2000	4000
RT60 (s)	0.24	0.23	0.24	0.23	0.23	0.25

TABLE I. Octave band reverberation time of the recording studio, averaged across the six microphone positions.

B. Binaural rendering and head tracking device

The "binauralization" of the resulting six-channel recording was made so as to give the impression of being at the center of the microphone array. For example, the signal from microphone 1 was processed using the HRTF for 0°, the signal from microphone 2 was processed using the HRTF for 60°, etc. The six resulting binaurally processed signals (one for each microphone) were then summed to generate the resulting left and right ear signals.

The rendering was carried out using the binaural engine *Bipan* (Baskind *et al.*, 2012) which uses anechoic measured HRTFs at either 15° or 5° azimuthal spacings. HRTFs are decomposed into minimum phase (for spectral cues) and pure delay (for ITD cues). Minimum phase transfer functions are modeled by infinite impulse response filters that are linearly interpolated every 1°. Thus, filters change every 1°, with a 1 ms cross-fade to smooth transitions between filters. ITD delays vary continuously as the subject moves his/her head using linear interpolations between the ITD of two consecutive known positions.

The head tracking was carried out using the open-source hardware/software solution *Hedrot*[3]. The tracking device, attached at all times to the subjects' headphones, consisted of an IMU GY-85 Sensor Module, with a Honeywell HMC5885L magnetometer, an Analog Devices ADXL345 accelerometer, and an Invensense ITG-3200 gyroscope. The head tracker was connected to the computer via a Teensy 3 USB board, and stimuli were updated in response to head movements at a rate of 300 Hz (3.3 ms). The total tracking system latency averaged 48.1 ms (SD = 5.3 ms).

Several non-individualized sets of HRTFs were interlaced instead of a single one, in order to investigate whether or not the impact of head tracking could change depending on the employed HRTF and also in order to minimize any HRTF learning effect. If only one HRTF set was used, it could be difficult to separate head tracking effects from those due to learning processes.

The HRTF sets chosen for the present experiment were n°1004, 1040, and 1077 from the publicly available LISTEN database (Warusfel, 2003). HRTF n°1040 was selected because several public demonstrations have suggested that this HRTF satisfied most subjects' judgments, and it was used by Nicol *et al.* (2016) for their comparative study of binauralized recording setups. The two other HRTFs were chosen on the basis of an informal test conducted by four of the authors, which suggested that perceptual differences between HRTFs n°1004, 1040,

and 1077 were substantial, thus providing a wide span of the perceptual range of HRTFs.

C. Azimuths

The tested azimuth positions spanned the horizontal plane at 30° intervals, at 0° elevation only. These azimuth directions correspond to the positions at which the signal obtained from the front microphone (microphone 1) was rendered. The rendered positions of the other microphones (microphones 2–6) were rotated accordingly: for example, a stimulus at +30° meant that the signal from microphone 1 was rendered at +30°, microphone 2 at +90°, microphone 3 at +150°, etc.

D. Reproduction Setup

The listening test took place in a double-walled soundproof booth at the Conservatory of Paris (background noise level ≈25 dB A). The lights were turned off in order to minimize the influence of any visual stimuli. The subject sat at the center of the room.

Stimuli were presented over headphones (Sennheiser HD 600). The sound pressure level was adjusted to ≈65 dB A (SLM, slow response) by placing the headphones on a dummy head (Neumann KU 100). Playback, interface, and data capture were controlled by software implemented in Max/MSP on a MacBook Pro computer connected to a RME Fireface 800 soundcard.

E. Subjects and Protocol

Ten subjects took part in the experiment (four women and six men, aged 22–57 years). They were financially compensated 60€ for their participation, none reported any known hearing loss. All subjects were professional sound engineers accustomed to listening to binaural content, yet none had experience with scientific listening tests.

Subjects were asked to either keep their heads as stationary as possible or to turn their heads back and forth between ±90°. The head tracker could either be active or inactive. Thus, subjects evaluated four different head tracking conditions:

SØ : static head orientation (no head movement), no head tracking.

ST : static head orientation, with head tracking.

MØ : with head movements, no head tracking.

MT : with head movements, with head tracking.

Subjects were requested to hold their heads in a natural upright position when listening to a stimulus. For conditions with head movements (**MØ** and **MT**), the presentation of the 8 s stimulus was divided into three phases:

1. 5.5 s of speech stimulus, during which subjects turned their heads in one full cycle first to the left (−90°) and then to the right (+90°) before returning to forward-facing (0°). All subjects were asked to make the same movements, as this ensured that they all received similar cues and that none provided differing results based on particularly efficient or ineffective choice of head movements. The form of controlled requested movements are similar to those proposed by Yairi *et al.* (2007) and Stitt *et al.* (2016a). The extent of motion was large and could be uncomfortable over the duration of the experiment. Subjects were seated on a swivel chair with the suggestion to carry out part of the motion through direct head movement and part through body/chair rotation to arrive at the target orientation.

2. 1 s silence. By the end of this silence, all head movements should be completed and subjects should be forward-facing again (0°), heads still.

3. 2.5 s of stimulus where subjects had to keep their heads stationary.

Grade	Reported externalization
0	The source is at the center of my head.
1	The source is not at the center of my head, but still in my head.
2	The source is at my ear, or on my skull.
3	The source is externalized but near the head.
4	The source is externalized and within my reach.
5	The source is externalized and remote.

TABLE II. Six-point scale used to report externalization.

After the final 2.5 s stimuli with head stationary, subjects reported to what extent the sound source was externalized using a six-point scale displayed on a computer screen (see Table II). The scale was inspired by several previous studies (Hartmann and Wittenberg, 1996; Kim and Choi, 2005; Kawaura *et al.*, 1991; Boyd *et al.*, 2012). Once subjects had given their answer, the next stimulus was automatically played.

In previous studies, subjects were to report to what extent a sound source had been externalized *while* they were moving their heads. In the present study, subjects reported to what extent a sound source was externalized *during the last* 2.5 s of the presentation, that is from the moment they were forward-facing and stationary again. In other words, subjects reported to what extent a sound source was externalized *after* they had moved their heads. Although this presented the risk that externalization may be high while subjects move their heads and then collapse once they stop moving, resulting

in a poorer reported externalization, this question protocol was preferred over those of previous studies because it enabled investigation of whether or not substantial improvements provided by head movements persist even though the subject has stopped moving his/her head. As mentioned in Section I D, the practical interest of head tracking would be severely reduced if it only improves externalization *while* the subjects move their heads, as it means that listeners would have to move continuously if they wish to listen to binaural content with optimal externalization. Moreover, ambiguous situations can arise with the protocols of previous studies: one could imagine a situation in which a subject, while moving his/her head during a presentation, would sometimes externalize the stimulus maximally (when the stimulus is at the extreme left for example), and sometimes would not (when the stimulus is directly in front for example). In that case, how should the subject respond, as the externalization question applies to the whole presentation of the stimulus? The protocol of the present study eliminates such ambiguities.

For the conditions without head movement (**SØ** and **ST**), the procedure was the same except that subjects were instructed to keep their heads still, looking straight ahead during the whole presentation of the stimulus.

For each of the four conditions [2 (head tracking yes/no) ×2 (head movements with/without)], there were 3 (HRTFs) ×12 (azimuths), resulting in a total of 36 trials grouped into a single block. Each block was repeated five times consecutively. Each condition took about 1 h to complete, and all subjects conducted the four conditions on four different days. The order of conditions was randomized and different for each subject. Within a condition, azimuth positions were presented in a randomized order that was different for each subject. The HRTF set always changed from one trial to another, thus minimizing potential HRTF learning effects.

III. RESULTS

A. Head Movements

During the test, head movements were recorded in order to verify how well the experimenters' instructions were followed by the subjects in all conditions.

1. Conditions without head movements: SØ and ST

Examination of data suggests that subjects were compliant with the experimenters' instructions. For conditions **SØ** and **ST**, the median amplitudes of movement (defined as the difference between the maximum and minimum angles over the course of a given trial) were 1.5° (inter-quartile range 1.7°) and 1.8° (inter-quartile range 2.5°) respectively.

The amplitude of movement was $\leq 1.5°$ for 39% of the trials during condition **ST**. According to Carlile and Leung (2016), data from several studies spanning 1971 to 2014 show that the minimum audible movement angle (MAMA) for wide band stimuli, defined as the minimum distance that a stimulus needs to be moved to be distinguished from a stimulus that is stationary, is $\geq 1.5°$ for durations of movement less than 200 ms, and then appears to asymptote at $\approx 1.5°$ for durations greater than 200 ms. Thus, it can be assumed that there were many trials during condition **ST** where the subjects' movements were too small to elicit any perceptible differences in spite of the active head tracking.

There were still trials during condition **ST** where the amplitudes of movement were larger and likely to provoke perceptible differences ($\geq 3°$ in 30% of the trials). Nevertheless, there was no substantial correlation observed between the amplitudes of movement and the externalization scores ($\rho = 0.091$, Spearman's rho). This implies that even the largest involuntary movements did not necessarily lead to more externalization.

Similarly, there was no substantial correlation between the amplitudes of movement and the externalization scores for condition **SØ** ($\rho = 0.009$).

2. Conditions with head movements: MØ and MT

For condition **MØ**, the median minimum and maximum head angles were −94° and 103° with inter-quartile ranges of 20° and 26° respectively. The first peak occurred at a median value of 2.0 s (inter-quartile range 0.36 s) and the second peak occurred at a median value of 4.3 s (inter-quartile range 0.56 s). The median duration of the movement was 5.6 s (inter-quartile range 0.60 s) and the median speed of head motion was 72°/s (inter-quartile range 16°/s).

For condition **MT**, the median minimum and maximum angles were −96° and 105° with inter-quartile ranges of 22° and 21° respectively. The first peak occurred at a median value of 2.0 s (inter-quartile range 0.38 s) and the second peak occurred at a median value of 4.3 s (inter-quartile range 0.66 s). The median duration of the movement was 5.7 s (inter-quartile range 0.76 s) and the median speed of head motion was 74°/s (inter-quartile range 17°/s).

Thus, the turns made by the subjects overshot the requested angular extents most of the time. However, these overshoots were relatively small, and similar overshoots were observed in Stitt *et al.* (2016a) and with some subjects in Brimijoin *et al.* (2013). Moreover, there was no correlation between the amplitudes of movement and the externalization scores for both conditions **MØ** and **MT**, which means that the variability of amplitudes of head movements was not large enough to have a substantial impact on externalization results. Similarly, there was no correlation between the speeds of motion and the externalization scores.

Thus, examination of head movement data suggests that subjects were reasonably compliant with the different head movement instructions for all four conditions.

B. Influence of the HRTF set

A Friedman test revealed that there was no significant difference among the externalization scores of the three HRTF sets ($p = 0.735$). An in-depth examination of the data found that externalization scores were indeed very similar from one HRTF set to another, independent of condition and subject. Subsequent results are therefore presented averaged across the three HRTF sets.

C. Influence of condition and azimuth

As expected from Section I B, examination of data revealed that results could greatly vary between azimuths: for lateral azimuths (±60°, ±90°, ±120°), externalization was high and differences between conditions were either small or null; for rear azimuths (±150°, 180°) and frontal azimuths (0°, ±30°), externalization was lower and differences between conditions were much more pronounced. It was thus decided to present the results for lateral azimuths (Section III C 1), rear azimuths (Section III C 2), and frontal azimuths (Section III C 3) separately.

Subsequent results were analyzed using Wilcoxon tests. When multiple pairwise tests were performed simultaneously, p-values were systematically adjusted using the Bonferroni correction.

1. Lateral azimuths: ±60°, ±90°, ±120°

A series of Wilcoxon tests reveals that results for conditions **SØ**, **ST**, and **MT** were not significantly different from each other for any of the lateral azimuths (p-values were always $\gg 0.05$), apart from one exception at azimuth 120° where externalization was significantly higher for condition **MT** than for condition **SØ** ($p = 0.006$). These results therefore show that, for lateral azimuths, head movements coupled with head tracking (**MT**) *did not enhance* externalization substantially compared to conditions without head movement (**SØ** and **ST**).

However, Wilcoxon tests show that head movements without head tracking (**MØ**) did result in a lower externalization compared to the other conditions at all lateral azimuths, apart from a few exceptions. For example, at azimuth −90°, the differences between **MØ** and the other conditions were not statistically significant.

Fig. 2 (left and center) details mean externalization scores with associated 95% confidence intervals obtained for each subject, condition by condition, over all lateral azimuths. For most subjects, externalization was quite high for conditions **SØ** and **ST** (≥ 3), thus providing little room for improvement when head movements and head tracking (**MT**) were added. The figure also shows that, although externalization scores were not dramatically low during condition **MØ** (≥ 2.5 for all subjects), they could be substantially lower compared to the other conditions for some subjects (MG, VL, and JP).

Fig. 2 (right) shows normalized externalization scores averaged across all subjects. Results were mean-normalized before averaging across subjects. Normalization of data was conducted so that each subject's mean score over all trials was equal to the global mean score (*i.e.* the mean score over all subjects and all trials). This removed any bias due to the between-subject variation offsets in overall externalization rate, and thus focused on the relative changes in externalization across conditions. The plot highlights the fact that scores for conditions **SØ**, **ST**, and **MT** were high and very similar. For condition **MØ**, the mean score was lower compared to the other conditions, however the difference was quite slight.

2. Rear azimuths: −150°, 180°, +150°

A series of Wilcoxon tests shows that, for each rear azimuth (−150°, 180°, +150°), externalization was significantly higher for condition **MT** than for any of the other conditions ($p \leq 0.01$), and externalization was significantly lower for condition **MØ** than for any of the other conditions ($p \leq 0.001$). However, there was no significant difference between conditions **SØ** and **ST** for any of the rear azimuths ($p \gg 0.05$).

Fig. 3 (left columns) details the results of each subject at azimuths ±150° and 180°.

At azimuths ±150°, mean externalization scores were already high during the conditions without head movement (**SØ** and **ST**) for four subjects (CB, DS, JP, and SM; ≥ 3.5), and externalization was not improved substantially, or even at all, when head movements and head tracking (**MT**) were added. For four subjects (HM, JB, MG, and VL), mean externalization scores were lower during the conditions without head movement, **SØ** and **ST** (≤ 3), and high mean externalization scores (≥ 3.5) were obtained only when head movements coupled with head tracking (**MT**) were added.

At azimuth 180°, mean externalization scores were already high for conditions without head movement (**SØ** and **ST**) for three subjects (CB, DS, and JP; ≥ 3.3 in most cases), and externalization was not improved substantially, or even at all, when head movements and head tracking (**MT**) were added. For five subjects (HM, JB, MG, SM, and VL), mean externalization scores could be quite low for the conditions without head movement, **SØ** and **ST** (≤ 2.5 in most trials). However, head movements coupled with head tracking (**MT**) enabled to improve externalization substantially, especially for three subjects (JB, MG, and VL; $\geq +1.5$ compared to the other conditions).

It is noted that in most cases, head movements coupled

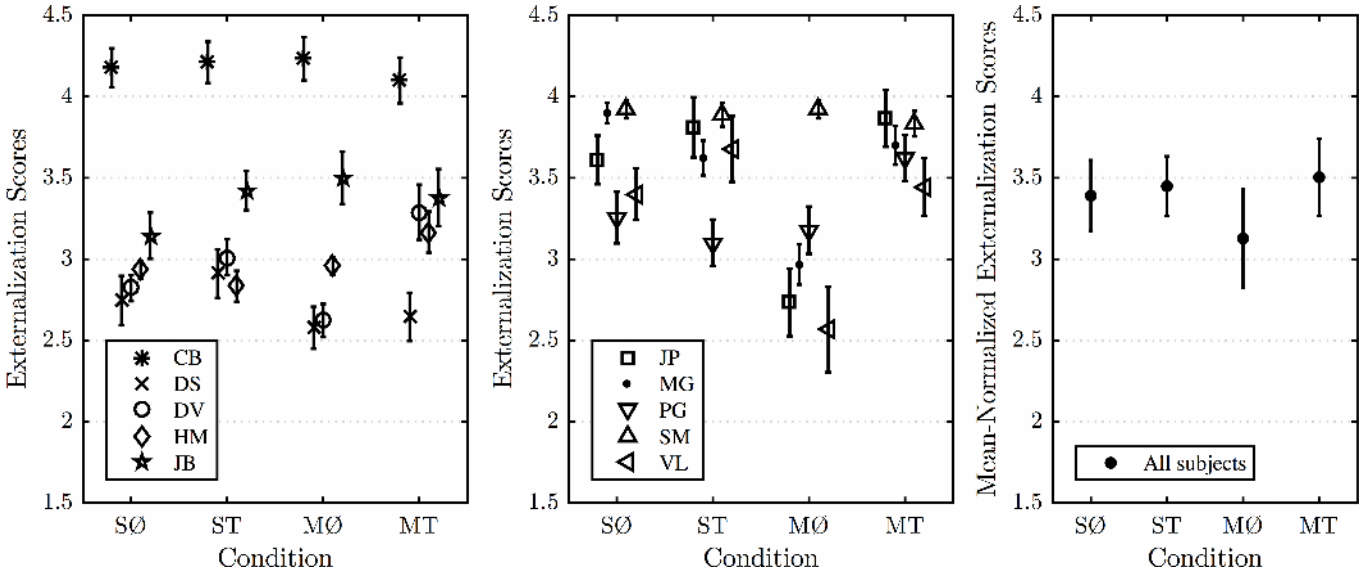

FIG. 2. Mean externalization scores with associated 95% confidence intervals obtained for each subject, condition by condition, over all lateral azimuths [±60°, ±90°, ±120°]. For visual clarity, individual results were split between two plots in alphabetical order (left and center). Mean-normalized externalization scores across all subjects (right).
SØ: no head movement, no head tracking. **ST**: no head movement, with head tracking.
MØ: with head movements, no head tracking. **MT**: with head movements, with head tracking.

with head tracking (**MT**) provided a substantial increase of externalization compared to head movements without head tracking (**MØ**) at both azimuths 180° and ±150°. The increase was especially high at azimuth 180°: from +1.2 to +3.3 for nine out of ten subjects.

Results at azimuths 180° and ±150° averaged across all subjects are presented in Fig. 3 (bottom-left). Results highlight the substantial improvement of externalization brought by head movements coupled with head tracking (**MT**) at azimuths 180°, which enabled to maintain a high global externalization, comparable with that of azimuths ±150° and lateral azimuths. The difference of externalization between conditions **MØ** and **MT** was especially pronounced for 180°. At azimuths ±150°, overall externalization was higher, minimizing the differences between conditions. At both azimuths ±150° and 180°, no clear advantage between conditions **SØ** and **ST** was observed.

3. Frontal azimuths: −30°, 0°, +30°

A series of Wilcoxon tests shows that, for each frontal azimuth (−30°, 0°, +30°), externalization was always significantly higher for condition **MT** than for any of the other conditions ($p \leq 0.01$), and externalization was always significantly lower for condition **MØ** than for any of the other conditions, apart from one exception: at azimuth 0°, the difference between **SØ** and **MØ** was not significant ($p \gg 0.05$). Again, there was no significant difference between conditions **SØ** and **ST** for any of the frontal azimuths ($p \gg 0.05$).

Fig. 3 (right columns) details the results obtained for each subject at azimuths 0° and ±30°. For conditions **SØ**, **ST**, and **MØ**, externalization was globally lower compared to that of the lateral and rear azimuths, especially at azimuth 0°, where individual mean scores were often very low (≤ 1). For condition **MT**, although head movements coupled with head tracking did not always allow for high scores (≤ 2.5 for eight out of ten subjects at azimuth 0°), they still enabled the observation of substantial improvements for most subjects:

- For four subjects (DS, HM, MG, and SM), even though the improvement brought by head movements coupled with head tracking (**MT**) could be moderate or even null at azimuth ±30° compared to other conditions, it was quite substantial at azimuth 0°: from +1.4 to +2.6 compared to condition **SØ**, from +0.8 to +1.6 compared to condition **ST**, and from +1.3 to +2.3 compared to condition **MØ**.

- For two subjects (CB and PG), substantial improvements compared to the other conditions could be observed at both azimuths 0° and ±30°. Improvements were especially pronounced at azimuth 0°: +1.6 and +1.7 respectively compared to condition **SØ**, +2.0 and +1.9 compared to condition **ST**, and +1.9 and +2.2 compared to condition **MØ**.

Mean-normalized externalization scores across subjects at azimuths 0° and ±30° are presented in Fig. 3 (bottom-right). The trend between conditions is very similar to that of the rear source positions: at azimuth 0°, substantial improvement in externalization could be observed for condition **MT** compared to other conditions;

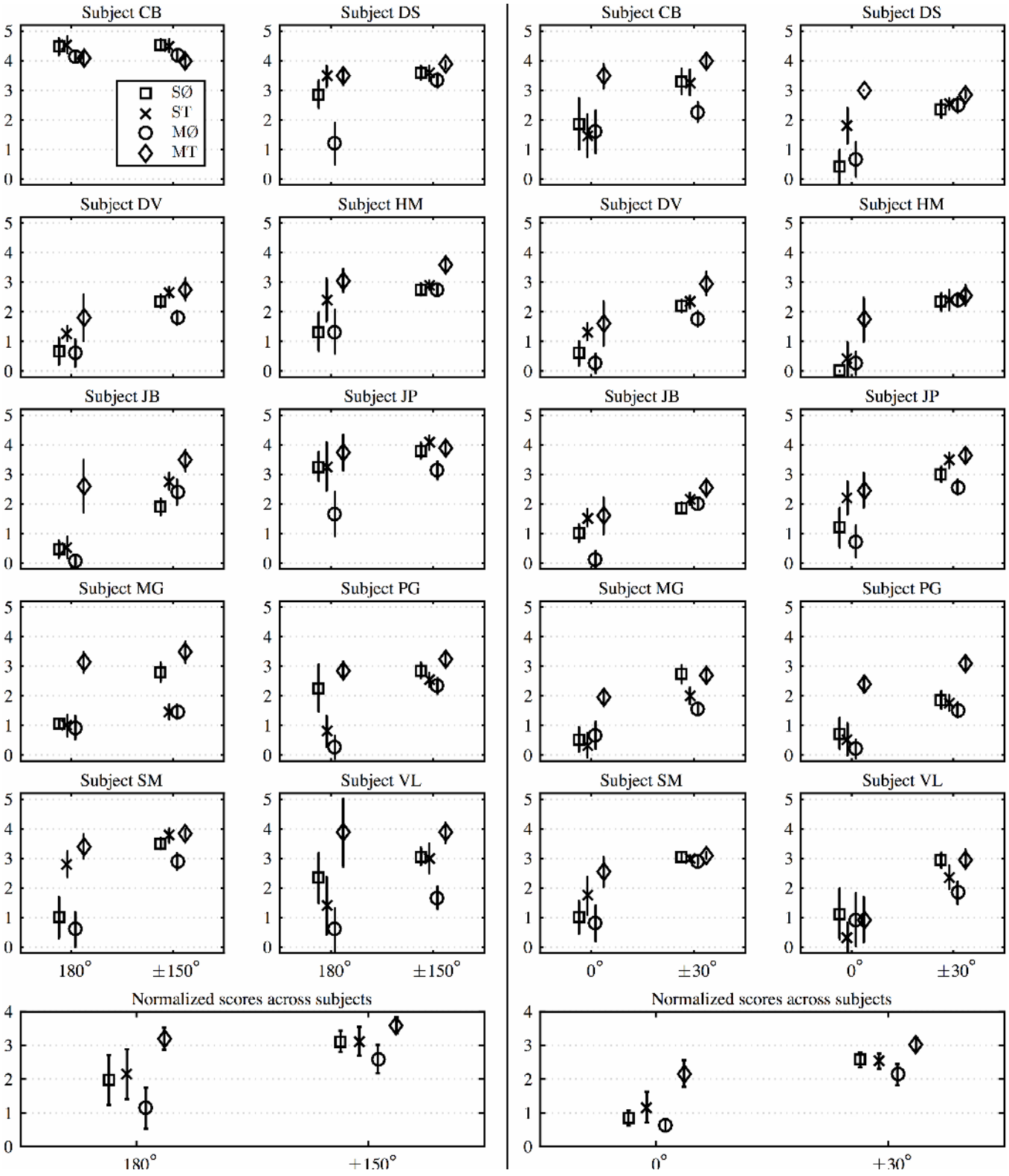

FIG. 3. Mean externalization scores with associated 95% confidence intervals obtained for each subject, condition by condition, at azimuths 180° and ±150° (left columns) and azimuths 0° and ±30° (right columns). The bottom figures show mean-normalized scores with associated 95% confidence intervals across all subjects.
□ : **SØ** (no head movement, no head tracking). × : **ST** (no head movement, with head tracking).
○ : **MØ** (with head movements, no head tracking). ◊ : **MT** (with head movements, with head tracking).

at azimuths ±30°, overall externalization was higher and differences between conditions were reduced; at both az-

imuths 0° and ±30°, no clear advantage between conditions **SØ** and **ST** was observed. Mean-normalized results highlight the fact that externalization was not high at azimuth 0° for condition **MT** (mean-normalized score = 2.2), even though it was still a substantial improvement compared to the dramatically low externalization observed for condition **SØ**, **ST**, and **MØ** (between 0.6 and 1.2).

4. Summary

Results can be summarized as follows:

- No convincing improvement of externalization was observed for condition **ST** compared to condition **SØ**.
- Apart from a few exceptions, head movements without head tracking (**MØ**) resulted in a lower externalization compared to the other conditions for lateral, rear, and frontal azimuths.
- For lateral azimuths, head movements coupled with head tracking (**MT**) did not enhance externalization significantly compared to the conditions without head movements (**SØ** and **ST**). Externalization was already reasonably high for conditions **SØ** and **ST**, therefore there was little room for improvement when head movements and head tracking (**MT**) were added.
- For the rear and frontal azimuths, head movements coupled with head tracking (**MT**) significantly improved externalization compared to conditions without head movements (**SØ** and **ST**). Substantial improvements were observed for most subjects, yet the magnitudes and the azimuths at which these improvements occurred varied greatly between subjects. For three out of ten subjects (HM, MG, and SM), substantial improvements were observed at both frontal and rear quadrants. For four subjects (CB, DV, DS, and PG) substantial improvements were mainly observed for the frontal quadrant. For two subjects (JB and VL), substantial improvements were mainly observed for the rear quadrant. For the remaining subject (JP), substantial improvements were found at the frontal quadrant compared to condition **SØ**, but not compared to condition **ST**.

Even though head movements coupled with head tracking (condition **MT**) substantially improved externalization, scores were not always high, especially for frontal sources. In contrast, Brimijoin *et al.* (2013) obtained high externalization with frontal sources and individualized HRTFs, even when subjects did not move their heads (**SØ** and **ST**). This comparison highlights the importance of correct HRTFs in the phenomenon of externalization.

IV. DISCUSSION

A. Comparisons with previous studies

In spite of comparable conditions (speech stimuli, reproduced at ear-level all around the subject, with non-individualized HRTFs), Begault *et al.* (2001) found that head movements coupled with head tracking did not significantly enhance externalization. Begault *et al.* used three different levels of reverberation: anechoic (no reverberation), early reflections, and early reflections coupled with late diffuse reverberation. The difference in outcomes with the present experiment could therefore be explained by differences in reverberation, as many studies have reported that externalization is strongly linked to the amount of reverberation (Begault, 1992; Durlach *et al.*, 1992; Plenge, 1974; Sakamoto *et al.*, 1976). Nevertheless, Begault *et al.* did not find a correlation between head tracking and reverberation, as head tracking was not observed to enhance externalization, independent of the level of reverberation. Other factors might thus explain the difference in outcomes. It could be due to the fact that the stimulus of the present study was longer (8 s instead of 2−3 s), thus giving subjects more time to take advantage of cues derived from head movements, and enabling them to make larger movements.

In Wenzel (1995), who also used non-individualized HRTFs, subjects were asked to provide numerical estimates of distance in inches (the distance scale was anchored by 0 inches for a sound at the center of the head and 4 inches for a sound located at the perimeter of the head). Results were then converted into externalization rate (defined as the percentage of time a stimulus was perceived outside the head, *i.e.* estimation > 4 inches) and averaged across all positions. Head movements coupled with head tracking (equivalent to condition **MT**) improved externalization rate from 74.5% to 83.5% (a 9% improvement) compared to a situation in which there was no head-tracking and subjects were instructed to keep their heads stationary (equivalent to condition **SØ**). The improvement seems rather moderate. However, if results of the present study are converted into externalization rate (defined as the percentage of time externalization score is ≥ 3) and averaged across all azimuths as shown in Fig. 4, the improvements brought by head movements coupled with head tracking (**MT**) compared to the conditions without head movements (**SØ** and **ST**) also seem moderate for most subjects. The mean-normalized averages across subjects show differences relative to **MT** of ≈ +12% compared to condition **SØ** and ≈ +11% compared to condition **ST**. Only when examining results azimuth by azimuth does one realize that, while the improvement was null for lateral azimuths, it could be quite substantial at some frontal and rear azimuths. For example, in terms of externalization rates, the increase for condition **MT** at azimuths 0° and 180° was on average ≈ +38% when compared to condition **SØ** and ≈ +31% when compared to condition **ST**.

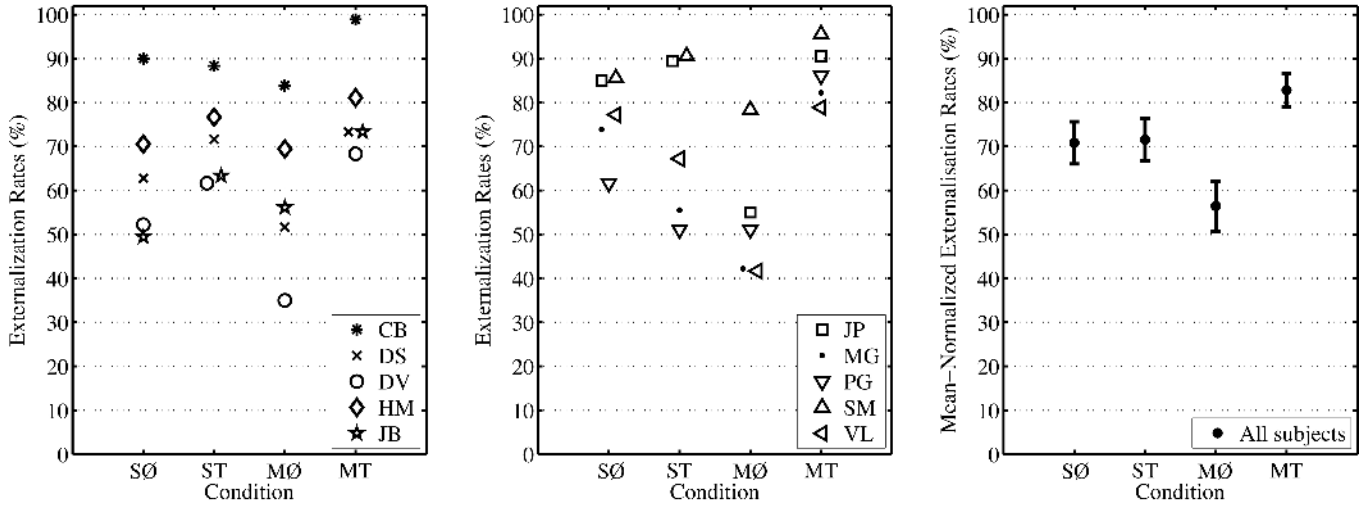

FIG. 4. Externalization rates obtained for each condition, over all azimuths. For visual clarity, individual results were split between two plots in alphabetical order (left and center). Mean-normalized externalization rates across subjects (right).
SØ: no head movement, no head tracking. **ST**: no head movement, with head tracking.
MØ: with head movements, no head tracking. **MT**: with head movements, with head tracking.

The present study therefore suggests that averaging externalization results across all tested positions should be avoided, because small or absent improvements observed at some positions tend to minimize much larger improvements observed at other positions. The fact that the effect of head tracking was not significant in Begault *et al.* (2001) might also be due to the fact that the analyses were conducted across all azimuths.

In Brimijoin *et al.* (2013), speech stimuli were reproduced in the horizontal plane for azimuths spanning ±25° (comparable to the frontal azimuths of the present study):

- With individual HRTFs, Brimijoin *et al.* observed that head movements coupled with head tracking did not improve externalization substantially compared to the conditions where the subject did not move his/her head, as externalization was already high even without head movements, precisely due to the fact that the binaural synthesis was individualized. However, externalization collapsed when subjects moved their heads without head tracking (**MØ**).

- With pure "head-absent" transfer functions (HATFs) and mixtures of individualized HRTFs and HATFs, Brimijoin *et al.* observed that head movements coupled with head tracking (**MT**) did improve externalization substantially compared to the other conditions.

These results and the present study therefore suggest that head movements coupled with head tracking might be more beneficial for externalization when the binaural synthesis is not individualized.

It is noted that subjects in the present study were not asked to report whether or not the stimulus was externalized *while* they were moving their heads, but *after* they had moved their heads. As expected, many subjects reported that stimuli were sometimes externalized while moving their heads, then internalized once they stopped moving their heads. It can therefore be hypothesized that the improvement brought by head movements and head tracking would have been even higher compared to previous studies if subjects had been asked to report externalization *while* moving their heads, as they were in Begault *et al.* (2001), Wenzel (1995) and Brimijoin *et al.* (2013).

B. No enhancement of externalization due to micro-movements of the head coupled with head tracking

No convincing improvement of externalization was observed for condition **ST** compared to condition **SØ**, for any azimuth. Hypothesis **H3**, that subjects when requested not to move their heads still make involuntary micro-movements that improve externalization if the head tracker is active, was, therefore, not verified in the present experiment. This agrees with Brimijoin *et al.* (2013), who also observed with individualized HRTFs and HATFs that head tracking was irrelevant when subjects were requested to keep their heads still.

This result could be explained by the fact that subjects' movements were too small in many trials to provoke any perceptible differences in spite of the active head tracking. As mentioned in Section III A 1, the amplitude of movement (defined as the difference between the maximum and minimum angles over the course of a given trial) was often inferior to the minimum audible move-

ment angles (MAMAs) reported in the literature.

Moreover, there was no substantial correlation between the amplitudes of movement and the externalization scores. Thus, even the largest involuntary movements ($\geq 3^\circ$ in 30% of the trials) did not necessarily lead to more externalization for condition **ST**. This finding supports hypothesis **H2**, that head movements need to be sufficiently large in order to have a substantial effect.

C. Head movements coupled with head tracking can be an effective way of providing more externalization

In the present study, head movements coupled with head tracking led to substantial improvements of externalization for most subjects, in support of hypothesis **H2**, and results suggest that such improvements can be observed with various non-individualized HRTF sets. Indeed, similar improvements were observed for the different HRTFs used in the present study, however further investigation with more HRTFs would be required to verify this result.

Moreover, the provided externalization appears to be robust. In condition **MT**, subjects were asked to report whether or not the stimulus was externalized *after* they had moved their heads. The fact that more externalization was obtained for that condition therefore shows that a stimulus, externalized by head movements and head tracking, can remain externalized even if the subject stops moving his/her head. Informal tests suggest that this externalization can persist as long as the stimulus remains the same (same voice, at the same azimuth, with the same HRTF).

D. Practical applications

The positive impact of head movements coupled with head tracking on externalization mostly occurred in zones that are critical in many applications.

The frontal quadrant (-30°, 0°, $+30^\circ$), for example, is especially important in situations such as virtual home theaters, because dialogs and on-screen sounds of a 5.1 mix are typically reproduced on the virtual center speaker located at 0°, while ambiance sounds and music are diffused on the front left and right virtual speakers located at $\pm 30^\circ$ (Toole, 2008). In other contexts such as teleconferencing or virtual reality, the frontal quadrant is often the most critical zone of interest as well, as it also represents the majority of the visual field of view.

The benefit of head tracking is further highlighted if one considers the fact that, in everyday life, a listener's head is rarely still and moves substantially in many situations (Kim *et al.*, 2013). In the present study, it can be observed in condition **MØ** that moving the head without head tracking considerably affected externalization in a negative way, at all azimuths, in support of hypothesis **H1**. Such a trend was also observed by Brimijoin *et al.* (2013), whether individualized HRTFs, "head-absent" transfer functions (HATFs) or mixtures of HRTFs and HATFs were used.

V. CONCLUSION

In the present study, a speech stimulus was presented over headphones with different source azimuths in the horizontal plane for ten experienced subjects using three interleaved sets of non-individualized HRTFs. The head tracker could either be active or inactive, and subjects could either be asked to rotate their heads or to keep as still as possible. Results show that:

- Head movements coupled with head tracking can enhance externalization substantially for frontal and rear sources compared to a situation where the listener does not move his/her head.
- Head movements coupled with head tracking can enhance externalization to an even further extent, and at all azimuths, compared to a situation where the listener moves his/her head without head tracking (a very common headphone listening scenario).

Results and comparisons with previous studies suggest that head movements may need to be sufficiently large in order to have a substantial impact. If this condition is met, then substantial improvements provided by head movements can be observed with most subjects. These improvements appear to be robust, as externalization persisted over time even though the subject has stopped moving his/her head.

ACKNOWLEDGMENTS

The authors would like to thank Alexis Baskind, Thibaut Carpentier, Vincent Koehl, Julian Palacino, Mathieu Paquier, Claire Voirin, Olivier Warusfel and all the subjects who took part in the subjective experiments. This work was funded in part by the French FUI project BiLi ("Binaural Listening", www.bili-project.org, FUI-AAP14).

APPENDIX: RECORDING SETUP

The following section provides additional details about the six-channel equal-segment microphone array that was used to record the stimulus.

One of the main issues with microphone arrays is microphone "leakage". For example, direct sound from a frontal sound source will be picked up, or "leaked", into the microphones dedicated to the rear directions and lead to confusions of localization due to the direct sound being perceived both front and back. One solution to reduce the perceived effect of this acoustic cross-talk is to use

directional microphones and spaced arrays instead of coincident or near-coincident arrays (Williams, 2005). This enables the rear microphones to capture direct sound with less intensity and greater delay, thus strengthening the precedence effect (Haas, 1949) directed towards the front microphone signal.

The recording system used in the present experiment enabled reduction of cross-talk effectively, as it used spaced cardioid microphones:

- Microphones 2 and 6 were at a distance of 0.95 m to the talker, whereas microphone 1 was at a distance of 0.50 m to the talker. There was therefore a ≈ 6 dB attenuation of the direct sound for microphones 2 and 6 compared to microphone 1, as the distance to the sound source was almost doubled. Moreover, due to the cardioid patterns of the microphones, the response was down by ≈ 6 dB as the talker was captured by microphones 2 and 6 with an angle of 93°. Thus, the total attenuation of the direct sound on microphones 2 and 6 was ≈ 12 dB compared to microphone 1.

- Microphones 3 and 5 were at a distance of 1.50 m to the talker, which corresponds to an attenuation of ≈ 10 dB. Moreover, the talker was captured with an angle of 135°, which means that the response was down by ≥ 12 dB. Thus, the total attenuation of the direct sound on microphones 3 and 5 was ≥ 22 dB compared to microphone 1.

- Microphone 4 was at a distance of 1.70 m to the talker, which corresponds to an attenuation of $\approx$ 11 dB compared to Microphone 1. Moreover, the talker was captured with an angle of 180°, which means that the level of direct sound was substantially reduced compared to microphone 1.

The retained spaced microphone array also enabled the direct sound from the talker to be delayed on microphones 2 to 6 compared to microphone 1: by about 1.3 ms for microphones 2 and 6, 2.9 ms for microphones 3 and 5, and 3.5 ms for microphone 4. In Blauert (1971), identical broadband (music and noise) signals were presented to subjects from the front and the rear simultaneously. Between the front and rear signal, a time delay could be set, and it was found that the direction of the sound sensation coincided with the angle of incidence of the first wavefront for delay times greater than about ± 550 µs. In Blauert (1997), for stereophonic loudspeakers radiating coherent signals, a delay of 1.1 ms was sufficient for the resultant phantom image to be localized at the position of the earlier loudspeaker.

In the present experiment, delays were larger, and they were reinforced by substantial differences of intensity. Thus, it can be assumed that cross-talk was reduced effectively and that the voice of the talker was perceived in the direction of the first-arriving sound, that is the direction at which the signal from microphone 1 was rendered. The relative attenuation from the cardioid patterns and the distances also enabled to make delay-and-add filtering (*i.e.*"comb filtering" due to outputs from several spaced microphones being summed) unimportant.

The six-channel equal-segment microphone array was also selected because equal segmentation of the sound field enables continuous and homogeneous sound field capture in the horizontal place (Williams, 1991), and because informal comparative studies of several microphone arrays with ten subjects suggested that this configuration provided the most natural audio scene when binauralized.

NOTES

[1]The four tracking conditions were not referred using the condition labels **SØ**, **ST**, **MØ** and **MT** in Brimijoin *et al.* (2013). These labels are proposed by the authors of the present study in order to simplify the presentation of subsequent results.

[2]Some early analysis of preliminary results of this study have been previously presented (Stitt *et al.*, 2016b).

[3]https://abaskind.github.io/hedrot/

Baskind, A., Carpentier, T., Noisternig, M., Warusfel, O., and Lyzwa, J. M. (**2012**). "Binaural and transaural spatialization techniques in multichannel 5.1 production", in *27th Tonmeistertagung, VDT International Convention*.

Begault, D. R. (**1992**). "Perceptual effects of synthetic reverberation on three-dimensional audio systems", J. Audio Eng. Soc. **40**, 895–904.

Begault, D. R. and Wenzel, E. M. (**1993**). "Headphone localization of speech", Hum. Fac. Erg. Soc. **35**, 361–376.

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.

Blauert, J. (**1971**). "Localization and the law of the first wavefront in the median plane", J. Acoust. Soc. Am. **50**, 466–470, doi: 10.1121/1.1912663.

Blauert, J. (**1997**). *Spatial hearing: the psychophysics of human sound localization*, 222–224 (MIT press, Cambridge).

Boyd, A. W., Whitmer, W. M., Soraghan, J. J., and Akeroyd, M. A. (**2012**). "Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers", J. Acoust. Soc. Am. **131**, 268–274, doi:10.1121/1.3687015.

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**, doi: 10.1371/journal.pone.0083068, e83068.

Carlile, S. and Leung, J. (**2016**). "The perception of auditory motion", Trends hear. **20**, 1–19, doi:10.1177/2331216516644254.

Durlach, N. I., Rigopulos, A., Pang, X. D., Woods, W. S., Kulkarni, A., Colburn, H. S., and Wenzel, E. M. (**1992**). "On the externalization of auditory images", Presence-Teleop. Virt. **1**, 251–257, doi:10.1162/pres.1992.1.2.251.

Haas, H. (**1949**). "The influence of a single echo on the audibility of speech", J. Audio Eng. Soc. **20**, 145–159, english translation (1972).

Hartmann, W. M. and Wittenberg, A. (**1996**). "On the externalization of sound images", J. Acoust. Soc. Am. **99**, 3678–3688, doi:10.1121/1.414965.

Katz, B. F. and Parseihian, G. (**2012**). "Perceptually based head-related transfer function database optimization", J. Acoust. Soc. Am. **131**, 99–105, doi:10.1121/1.3672641.

Kawaura, J. I., Suzuki, Y., Asano, F., and Sone, T. (**1991**). "Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear", J. Acoust. Soc. Jpn. **12**, 203–216, doi:10.1250/ast.12.203.

Kim, C., Mason, R., and Brookes, T. (**2013**). "Head movements made by listeners in experimental and real-life listening activities", J. Audio Eng. Soc. **61**, 425–438.

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, doi: 10.1121/1.1921548.

König, G. and Sussmann, W. (**1955**). "Zum richtungshören in der median-sagittal-ebene [on directionnal hearing in the medial-saggital planes]", European Archives of Oto-Rhino-Laryngology **167**, 303–307, doi:10.1007/BF02107754.

Laws, P. and Platte, H. J. (**1975**). "Spezielle experimente zur kopfbezogenen stereophonie [some experiments in head-related stereophony]", in *Fortschritte der Akustik, DAGA 75', Physik-Verlag, Weinheim*, 365–368.

Loomis, J. M., Hebert, C., and Cicinelli, J. G. (**1990**). "Active localization of virtual sounds", J. Acoust. Soc. Am. **88**, 1757–1764, doi:10.1121/1.400250.

Martin, R. L., McAnally, K. I., and Senova, M. A. (**2001**). "Free-field equivalent localization of virtual audio", J. Audio Eng. Soc. **49**, 14–22.

Mendonça, C., Campos, G., Dias, P., Vieira, J., Ferreira, J. P., and Santos, J. A. (**2012**). "On the improvement of localization accuracy with non-individualized HRTF-based sounds", J. Audio Eng. Soc. **60**, 821–830.

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.

Nicol, R., Gros, L., Colomes, C., and Messonnier, J.-C. (**2016**). "Etude comparative du rendu de différentes techniques de prise de son spatialisée après binauralisation [comparative study of several spatial audio recording setups after binauralization]", in *Proceedings of Acoustics 2016 Conference* (Le Mans, France).

Noble, W. (**1987**). "Auditory localization in the vertical plane: Accuracy and constraint on bodily movement", J. Acoust. Soc. Am. **82**, 1631–1636, doi:10.1121/1.395154.

Perrett, S. and Noble, W. (**1997**). "The contribution of head motion cues to localization of low-pass noise", Percept. Psychophys. **59**, 1018–1026, doi:10.3758/BF03205517.

Plenge, G. (**1974**). "On the differences between localization and lateralization", J. Acoust. Soc. Am. **56**, 944–951, doi: 10.1121/1.1903353.

Politis, A., Laitinen, M. V., Ahonen, J., and Pulkki, V. (**2015**). "Parametric spatial audio processing of spaced microphone array recordings for multichannel reproduction", J. Audio Eng. Soc. **63**, 216–227, doi:10.17743/jaes.2015.0015.

Rébillat, M., Boutillon, X., Corteel, E., and Katz, B. F. (**2012**). "Audio, visual, and audio-visual egocentric distance perception by moving participants in virtual environments", ACM Transactions on Applied Perception, Association for Computing Machinery **9**, 1–17, doi:10.1145/2355598.2355602.

Sakamoto, N., Gotoh, T., and Kimura, Y. (**1976**). "On "out-of-head localization" in headphone listening", J. Audio Eng. Soc. **24**, 710–716.

Simon, L., Zacharov, N., and Katz, B. F. (**2016**). "Perceptual attributes for the comparison of Head-Related Transfer Functions", J. Acoust. Soc. Am. **140**, 3623–3632, doi: 10.1121/1.4966115.

Stitt, P., Hendrickx, E., Messonnier, J. C., and Katz, B. (**2016**a). "The influence of head tracking latency on binaural rendering in simple and complex sound scenes", in *Proceedings of the 140th Convention of the Audio Engineering Society*, 9591:1–8, paper no. 9591.

Stitt, P., Hendrickx, E., Messonnier, J.-C., and Katz, B. F. (**2016**b). "The role of head tracking in binaural rendering", in *Tonmeistertagung TMT*, 350–355 (Verband Deutscher Tonmeister, Cologne).

Toole, F. E. (**2008**). *Sound Reproduction: Loudspeakers and rooms*, 98–116 (Focal Press, Burlington), doi:10.1016/B978-0-240-52009-4.50005-8.

Warusfel, O. (**2003**). "LISTEN HRTF database", URL `http://recherche.ircam.fr/equipes/salles/listen/`, last visited 7-Feb-2017.

Wenzel, E. M. (**1995**). "The relative contribution of interaural time and magnitude cues to dynamic sound localization", in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 80–83, doi:10.1109/ASPAA.1995.482963.

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, doi: doi:10.1121/1.407089.

Wersényi, G. (**2009**). "Effect of emulated head-tracking for reducing localization errors in virtual audio simulation", in *IEEE transactions on audio, speech, and language processing*, volume 17, 247–252, doi:10.1109/TASL.2008.2006720.

Wightman, F. L. and Kistler, D. J. (**1989**). "Headphone simulation of free-field listening. II: Psychophysical validation", J. Acoust. Soc. Am. **85**, 868–878, doi:10.1121/1.397558.

Wightman, F. L. and Kistler, D. J. (**1999**). "Resolution of front-back ambiguity in spatial hearing by listener and source movement", J. Acoust. Soc. Am. **105**, 2841–2853, doi: 10.1121/1.426899.

Williams, M. (**1991**). "Microphone arrays for natural multiphony", in *Proceedings of the 91st Convention of the Audio Engineering Society*, paper no. 3157.

Williams, M. (**2005**). "The whys and wherefores of microphone array crosstalk in multichannel microphone array design", in *Proceedings of the 118st Convention of the Audio Engineering Society*, paper no. 6373.

Yairi, S., Iwaya, Y., and Suzuki, Y. (**2007**). "Estimation of detection threshold of system latency of virtual auditory display", Appl. Acoust. **68**, 851–863, doi:10.1016/j.apacoust.2006.12.005.