

The benefit of head orientation to speech intelligibility in noise

Jacques A. Grange^{a)} and John F. Culling

School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff CF103AT, United Kingdom

(Received 15 July 2015; revised 10 November 2015; accepted 25 January 2016; published online 9 February 2016)

Spatial release from masking is traditionally measured with speech in front. The effect of head-orientation with respect to the speech direction has rarely been studied. Speech-reception thresholds (SRTs) were measured for eight head orientations and four spatial configurations. Benefits of head orientation away from the speech source of up to 8 dB were measured. These correlated with predictions of a model based on better-ear listening and binaural unmasking ($r = 0.96$). Use of spontaneous head orientations was measured when listeners attended to long speech clips of gradually diminishing speech-to-noise ratio in a sound-deadened room. Speech was presented from the loud-speaker that initially faced the listener and noise from one of four other locations. In an undirected paradigm, listeners spontaneously turned their heads away from the speech in 56% of trials. When instructed to rotate their heads in the diminishing speech-to-noise ratio, all listeners turned away from the speech and reached head orientations associated with lower SRTs. Head orientation may prove valuable for hearing-impaired listeners. © 2016 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4941655>]

[MAH]

Pages: 703–712

I. INTRODUCTION

Spatial release from masking (SRM) is the reduction in speech-reception threshold (SRT) obtained from spatially separating one or several interferers from the target speech. SRM has been extensively studied to better understand the spatially dependent part of the cocktail party problem (Cherry, 1953) in normally hearing (NH) and hearing impaired (HI) listeners (Bronkhorst and Plomp, 1988, 1992; Dirks and Wilson, 1969; Peissig and Kollmeier, 1997; Plomp and Mimpen, 1981; Plomp, 1976). Other cues affecting the masking of speech include the modulation of interfering sounds, the relative fundamental frequencies (F0s) of target and interfering voices and the linguistic content of interfering speech (Miller, 1947; Hawley *et al.*, 2004). SRM is measured in most studies with listeners facing the speech; yet, Kock (1950) highlighted the substantial head-orientation benefit (HOB) of turning one's head away from the speech source. If head orientation away from the speech can provide a significant HOB, both NH and HI listeners could benefit from exploiting it.

Kock (1950) was the first to map out thresholds of speech intelligibility in noise as a function of head orientation away from the speech. The effect appears not to have been investigated since. Kock found that speech understanding was poorer when facing the speech source and also when the head orientation was such that speech and noise directions lay at the same angle away from the interaural axis (i.e., on the same cone of confusion). Figure 1 shows that this observation is qualitatively consistent with the predictions of the Jelfs *et al.* (2011) model of SRM. The model predicts effective target-to-interferer ratios (eTIRs), improvements in eTIR corresponding with reductions in SRT and improvements in SRM. The

model is based on the additive combination of better-ear listening and binaural unmasking. Figure 1 plots eTIR predictions as a function of head orientation and for four different masker separations. The initial directions of target and masker are indicated by subscripts (e.g., target at 0° and masker at 180° is denoted as T_0M_{180}). Moving the masker away from 180° moves the minima in the eTIR pattern. Sharp eTIR minima (arrows) occur when the head orientation is such that both target and masker lie on the same cone of confusion, whilst eTIR maxima can be reached by rotating the head between target and masker azimuths.

Despite the early findings from Kock (1950), both Bronkhorst and Plomp (1990) and Koehnke and Besing (1996) proposed that clinical testing of binaural hearing should be conducted with the listener facing the speech with noise either collocated or separated from the speech by 90° . Facing the speech was considered a more natural listening attitude (Plomp, 1986). The selected spatial configurations thereafter became a standard for most studies and clinical tests. Yet, they present two limitations: first, as was observed much earlier by Kock (1950), head orientation away from the speaker could lead to much improved SRM; second, both Müller (1992) and Peissig and Kollmeier (1997) found that SRM was reduced at 90° , compared to neighboring noise separations. A 90° noise azimuth places the ear contralateral to the noise in a noise bright spot due to the noise wave-front wrapping around both sides of the head and constructively interfering at that ear (Duda and Martens, 1998). Because the bright spot coincides with the contralateral ear only for a narrow range of source azimuths, small changes in head orientation can have large effects on SRT. Thus, in these standard configurations, not only is SRM not optimal, but the SRTs are also more variable.

Whilst head orientation away from the speaker can provide speech-in noise intelligibility improvement, facing, or

^{a)}Electronic mail: grangeja@cardiff.ac.uk

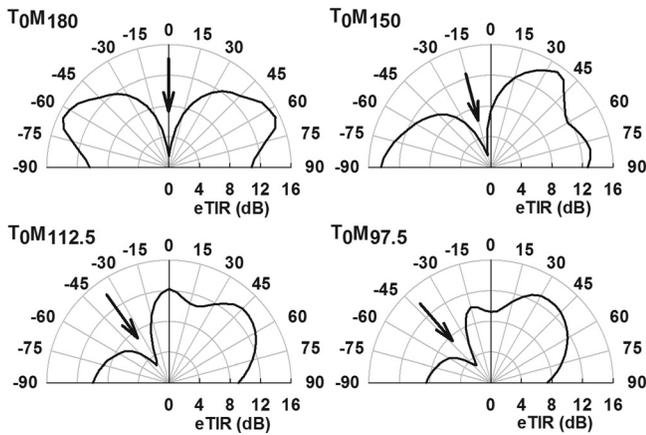


FIG. 1. Predicted eTIR (scaled as SRM) in the four spatial configurations used in the present study (T_0M_{180} , T_0M_{150} , $T_0M_{112.5}$, and $T_0M_{97.5}$). SRM is never at its highest at zero degrees (facing the speech) and it dips when the head orientation brings both sources onto the same cone of confusion (arrows).

at least looking at the speaker is a social norm (Kendon, 1967). All sighted listeners, and in particular, HI listeners are reliant on lip-reading when the speaker is visible and listening conditions are challenging (Schorr *et al.*, 2005; Sumbly and Pollack, 1954). A trade-off may be expected between the benefits of head orientation and that of lip-reading. However, as long as head orientation away from the speaker does not impede lip-reading, head-orientation and lip-reading benefits could potentially be combined. The remit of this paper is limited to audio-only presentation. How head orientation and lip-reading could be combined will be the object of future work.

Brimijoin *et al.* (2012) may have been the first to measure head orientation during a free-head listening task. The authors' aim was to establish what strategy listeners spontaneously employed when presented with short sentences from random directions masked by a single speech-shaped noise masker close to intelligibility threshold. They tested asymmetrically HI participants (>16 dB asymmetry), speculating that they may be more likely to make use of head movements than NH listeners. The data from Brimijoin *et al.* (2012) suggested that their listeners aimed for maximizing the target speech level, a strategy somewhat simpler, but less effective than aiming for optimum signal-to-noise ratio (SNR). However, the study by Brimijoin *et al.* (2012) presents several issues. First, there is no equivalent study in NH listeners that can be used as a benchmark. Second, the authors claimed that the behavior investigated was undirected and therefore indicative of natural strategies, but the listeners wore a tracking device that could have led them to think that the researchers were interested in head movements. Also, listeners "were told that the chair on which they were sitting could rotate and they should feel free to turn if they liked." This could be interpreted as an instruction or an implied direction. We aimed to address these limitations. The present study begins by testing the behavior of NH listeners with the aim of developing a robust paradigm for later experiments with HI patients. We endeavored not to say,

anything to participants that could give them any clue whatsoever that head orientation is a key focus of our experiment, and we measured their head orientation covertly, using overhead video recording.

Our study investigated whether normally hearing listeners can and do benefit from adopting appropriate head orientations away from the speech direction. It thus had two objectives, first, to confirm the HOB predictions of the Jelfs *et al.* (2011) model through fixed-head measures of SRT, and second, to test whether listeners spontaneously make use of the optimum head orientations predicted by the model. In addition to these primary questions, there was also the secondary issue of what strategies, if any, are used by listeners for finding the optimum head orientation(s). If listeners are able to exploit the potential benefits of head rotation, they could achieve this in at least four different ways. First, they may use gradual head turns to scan for speech intelligibility improvements. Second, they may localize the sources that are present and then predict, from their analysis of the auditory scene, the head orientation for optimal SNR. Third, as Brimijoin *et al.* (2012) suggested, listeners may focus on the target alone and optimize target level at one ear rather than SNR. Such a strategy would be unaffected by masker position; the optimal head orientation with respect to the target will always be at $\pm 60^\circ$ with respect to the speech target in this case. Finally, perception of the masker in one hemifield may influence the listener to turn away from the masker, which the model predicts would worsen their speech intelligibility.

II. METHODS

A. Experimental approach

The masker locations with respect to the target were originally selected to help differentiate between the listening strategies listed in the previous paragraph. The chosen angles were 180° , 150° , 112.5° , and 97.5° . The effects of head orientation were predicted using the Jelfs *et al.* (2011) model with the Gardner and Martin (1995) head-related impulse responses from KEMAR. The predictions for each spatial configuration are shown in Fig. 1. The largest HOB, of ~ 14 dB, is predicted when the target is initially in front and the masker initially directly behind.

In contrast with Brimijoin *et al.* (2012), we opted to separate behavioral (free-head) and threshold (fixed-head) measurements. In order to give listeners ample opportunity to make use of head orientation in the free-head paradigm, we presented long clips with gradually diminishing SNR. At the start of a trial, the SNR was high, such that listeners could follow the content of the clip with ease. The expectation was that as SNR approached the speech-facing SRT, listeners would increasingly be motivated to make use of head orientation. We also ensured that we would not compromise the undirected nature of the behavioral experiment. Care was taken to ensure that no reference (verbal or written) was made to head orientation until the free-head observations were completed. Each participant started their session with undirected, free-head listening tests. A complete set of SRT measurements followed the first behavioral experiment.

Finally, if time allowed, a second free-head experiment was run, this time in a directed manner.

B. Materials and methods

1. Participants

Sixteen participants were recruited from the Cardiff University undergraduate population (14 via the School of Psychology Experimental Management System) as well as the general population. They ranged from 19 to 50 years old, and 14 were less than 22 years old; the remaining two were 35 and 50 years old. All reported normal hearing but were not specifically screened for hearing loss. Each subject was tested over a single session of 3.5 h (including breaks) and was remunerated for their time. All were treated in accordance with the rules of our institution's Ethics Committee, with written briefing, consent, and debriefing forms supplied.

2. Laboratory setup

A 3.2 m × 4.3 m sound-deadened room was equipped with a 3-m diameter semi-circular array of 24 Cambridge Audio Minx speakers fitted 1.3 m above the floor. The speakers were driven by four Auna 6-channel solid-state amplifiers, themselves driven by a Motu 24-channel digital-to-analogue converter. All stimuli were controlled by MATLAB (Mathworks) bespoke programs, making use of the Playrec toolbox (Humphrey 2008–2014). Each channel of the audio chain was judged to be sufficiently consistent for our purpose in level and spectral response via acquisition of impulse responses and comparison of corresponding excitation patterns (Moore and Glasberg, 1983). The reverberation time (RT60) of the sound-deadened room was measured from these impulse responses to be ~60 ms. An adjustable swivel chair was positioned in the room such that once a participant was sat on the chair, their head would be at the centre of the speaker array regardless of chair rotation. Control of the experiments was achieved from a computer station in the room. A Microsoft Lifecam 5000 video camera was fitted on the ceiling exactly above the listener's head so that covert video recording could be made of the listener's head orientation.

3. Fixed-head SRT task: Material and protocol

SRT measurement used the IEEE sentence corpus (Rothausler *et al.*, 1969) in a one-up/one-down adaptive threshold method developed from that of Plomp and Mimpen (1979) and previously used in Lavandier and Culling (2010). The target speech was presented from the 0° loudspeaker and continuous speech-shaped noise was presented simultaneously with the speech from the 180°, 150°, 112.5°, or 97.5° azimuth loudspeakers to create the four selected spatial configurations. The noise sound level was kept constant at 70 dB sound pressure level (SPL) measured by a hand-held sound level meter at the listener position. The listener was required to repeat as much of a presented sentence as they could and the experimenter input the number of correctly recalled key words. The staircase started at a

low (–20 dB) SNR, so that even in the most favorable condition, none of the listeners could understand the first sentence. The first sentence was presented again with speech level increased in 4-dB steps until the listener correctly repeated at least three of the key words. From then on the adaptive phase started: the sentence was changed every trial and SNR was increased or decreased by 2 dB when the listener repeated correctly less or more than half of the keywords, respectively. The last eight SNRs computed were then averaged to calculate the SRT. The standard error across the final eight trials was computed to confirm the reliability of the averaged SRT. In order to ensure that the participants would remain still and facing the correct orientation for the duration of each trial, they were asked to face their own image in an appropriately positioned mirror and ensure symmetry of their own reflection.

Eight azimuthal orientations of the head were selected within each spatial configuration to construct a partial map of SRTs around the speech-facing orientation. The exact azimuths chosen within each spatial configuration aimed at confirming the predicted SRM maxima and minima (Fig. 1) surrounding the speech facing orientation. Selected head azimuths at T_0M_{180} were 0°, 15°, 30°, 52.5°, 60°, 67.5°, 75°, and 90°; at T_0M_{150} , –45°, –15°, 0°, 22.5°, 30°, 37.5°, 45°, and 60°; at $T_0M_{112.5}$, –30°, –22.5°, –15°, –7.5°, 0, 7.5°, and 30°; and at $T_0M_{97.5}$, –15°, –7.5°, 0°, 7.5°, 15°, 30°, 45°, and 60°. Target speech was from the Massachusetts Institute of Technology recordings of voices DA or CW. A total of 32 lists were used to cover all combinations of head orientations and spatial configurations, 16 lists from each voice. The speech material was kept in the same order for all participants. Trials were grouped in four spatial-configuration blocks, each block covering eight head orientations. The spatial-configuration blocks and the head orientations within each block were simultaneously rotated for each new participant, such that each voice was used equally across each of the 32 conditions.

4. Free-head task: Materials and protocol

The material used for this experiment consisted of four 4-min-long speech clips. They were speeches by President Obama obtained from the White House official website. All four clips exhibited consistency of speaker, speech flow, complexity and level. They talked of US internal or international affairs in a manner that was easy to follow. For each participant, each of these clips was allocated to one of the four spatial configurations. The voice from each clip was utilized to synthesize masking noise matched in long-term frequency spectrum to that voice. This speech-shaped noise was created using a 512-point finite-impulse-response filter that was based on the calculated excitation pattern of the speech material (Moore and Glasberg, 1983). The target speech and fixed-level speech-shaped noise (70 dB SPL) were simultaneously presented to the listener in each of the spatial configurations described above. The speech level was initially set at 7.5 dB SNR (at source), such that the speech would initially be easily understandable. Speech level (and hence SNR) was steadily decreased at a rate of 7.5 dB per minute, in such a way that it would reach the listener's

speech reception threshold if the listener kept facing the speech about a third of the way through the clip and so ensuring no listener would reach the end of a clip. Gradual speech level diminution was used as a means to motivate the listener to turn their head when the SNR became challenging. The listeners were instructed as follows: “Please do whatever you would normally do in a social situation to understand the speech for as long as possible and simply say, ‘STOP’ when you have lost track of the speech.” The only restrictions given to the listeners and clearly stipulated as “only restrictions” were to “please keep the chair central in the room, remain seated, keep your back against the chair’s back rest and keep your arms resting on your lap or on the arm rests during the task.” This ensured that the listener’s head remained in the centre of the speaker array and that listeners not be tempted to use their hands to block noise or reflect sound into their ear. Listeners were further motivated to perform the task studiously by being told they would be quizzed on the content of the clip they listened to. The time at which listeners flagged losing track of the speech would subsequently allow us to determine a subjective measure of SRT for their final head orientation. Finally, listeners were led to face the speech when the clip started, simply by being told which loudspeaker the speech would come from. They were not instructed to face that speaker. At the end of the undirected runs, listeners were asked whether they had interpreted the instructions given as an implied suggestion to make use of head orientation.

The video recordings were post-processed blind in a semi-automated procedure that made use of the MATLAB mouse pointer function. Over two passes, an operator tracked with the pointer the locations of the centre of the listener’s head and the then the listener’s nose. The two sets of coordinates obtained were combined to extract the listener’s head orientation with respect to the target direction. This method was found to be accurate within $\pm 5^\circ$, which was judged satisfactory for our purpose.

Having acquired the undirected behavioral data, we then informed the listener that head orientation might be beneficial and, where time allowed, repeated the free-head test after completion of the SRT runs. Listeners were then told the following: “We were interested in the first experiment to see what head orientation strategies you would naturally adopt. Please repeat the first experiment, this time with the knowledge that you might understand the speech for much longer if you orient your head away from the speech direction. You might experience head-orientation benefits more one way rather than the other or equally either way or none at all.” The rest of the instructions remained the same as for the first behavioral experiment. The same speech material was used.

Because our main interest was to find out whether people would naturally rotate their heads and what their final head position would be, rotation of configuration against material was not judged essential for the free-head test. However both spatial configuration and their associated speech clip were rotated across participants.

Since covert video recording was employed, listeners were given the opportunity at the end of the experiment to have these recordings immediately destroyed if they so

wished. None of them took up this offer. Given the undirected nature of the first experiment, participants were urged at the end of their session to not discuss with their peers the object of our research. Specifically, they were asked to not divulge our interest in head orientation.

III. RESULTS

A. SRTs

The reliability of the SRT measurement was assessed by computing the standard error across the final 8 SNRs. The mean (across all 512 runs) of such standard error was 0.67 dB. This provided a satisfactory degree of confidence in our SRT measurements. The two older (35 and 50 years old) participants’ data fell well within the younger adults’ range, both in terms of SRT and in terms of HOB. Their data were therefore retained. SRTs were averaged across all 16 participants in each of the 32 conditions. The standard error of the SRTs (across-participant means) did not exceed 0.56 dB and averaged 0.41 dB across conditions. Figure 2 shows the model fit to the observed data for each spatial configuration.

Binaural room impulse responses (BRIRs) acquired in the test room were used to compare the SRT outcomes to model predictions. BRIRs were acquired with a B&K Head and Torso Simulator (HATS) placed on the swivel chair. BRIRs were acquired by rotating the chair to have the HATS face between -90° and $+90^\circ$ azimuth every 7.5° . In order to compare the pattern of SRTs with the model output, in terms of eTIR, the SRTs were inverted and their mean equalized to that of the eTIRs produced by the model across the set of spatial configurations (Fig. 2). The figure is scaled in terms of SRM by adjusting the zero point such that the model predicts 0 dB for a collocated target and interferer. Disparity between data and predictions was typically within less than 1 dB. A significant correlation was found between SRT data and predictions [$r = 0.96$, $t(30) = 18.5$, $p < 0.001$], with a regression slope of 0.82, indicating somewhat less variation in observed than predicted SRT. The only exceptions to data fitting the model within 1 dB were typically found where the

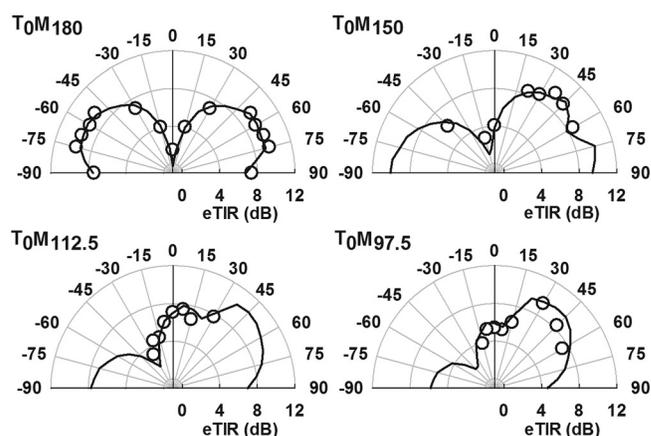


FIG. 2. Predicted eTIR (lines, scaled as SRM) and SRT data converted to SRM (circles) averaged over 16 participants for all spatial configurations (T_{0M180} , T_{0M150} , $T_{0M112.5}$, and $T_{0M97.5}$). The SRT data are inverted and their mean (across all 32 conditions) equalized to that of SRM predictions to enable comparison of data and prediction patterns across head orientations.

sharpest slope in benefit per degree of head rotation was predicted. There, the discrepancy was still less than 1.7 dB. The poorer fit at those points could be attributed to inaccuracy in listeners' head positioning during the SRT task, because only a slight deviation from the desired head orientations could give rise to a substantial change in SRM. It proved difficult for listeners to maintain a fixed and correct head orientation whilst focusing on the listening task at hand. Head orientation was not measured during the SRT runs. From time to time, the experimenter reminded participants, if they visibly deviated from the correct orientation, to check the symmetry of their own reflection in the mirror. It is plausible that listeners may have deviated at times by as much as 5° from the correct head orientation, which would account for most of the largest deviations from prediction.

An analysis of variance (ANOVA) on SRT for each spatial configuration found a significant effect of head orientation [$F(6,90) > 15.4$, $p < 0.001$]. A comparison of speech-facing SRTs revealed, as seen elsewhere, a significant effect of masker separation [$F(3,45) = 18.0$, $p < 0.001$], with means ranging from -10.8 dB (at T_0M_{180}) to -16.4 dB (at $T_0M_{112.5}$).

An effect of reverberation noticeable in the BRIR predictions is that the three asymmetric configurations showed broadly the same trend. The clear definition of minima and

maxima close to the speech-facing orientation found in HRIR predictions in Fig. 1 is very much dampened.

B. Free-head task

1. Undirected head movements

All listeners confirmed that they had not interpreted instructions as an implied suggestion to make use of head orientation. Most had assumed that our study focused on the effect of masker separation on their performance. Of the 64 undirected trials that were planned, data from 63 were obtained as one of the participants did not complete a full set of trials. The most significant finding was that in 56% of the trials (35 of 63) listeners spontaneously moved their head more than 10° away from the speech direction in response to the speech becoming increasingly difficult to follow.

In the T_0M_{180} configuration, a symmetrical benefit was predicted for a rotation of the head either way. The left panels of Fig. 3 show example time plots of undirected head orientations (solid lines) adopted by a subset of participants at T_0M_{180} . The open circles at the end of each track correspond to the clip time at which listeners flagged losing track of the speech. The SNRs reached at these points can be regarded as subjective measures of SRT achieved at the final head orientation, but

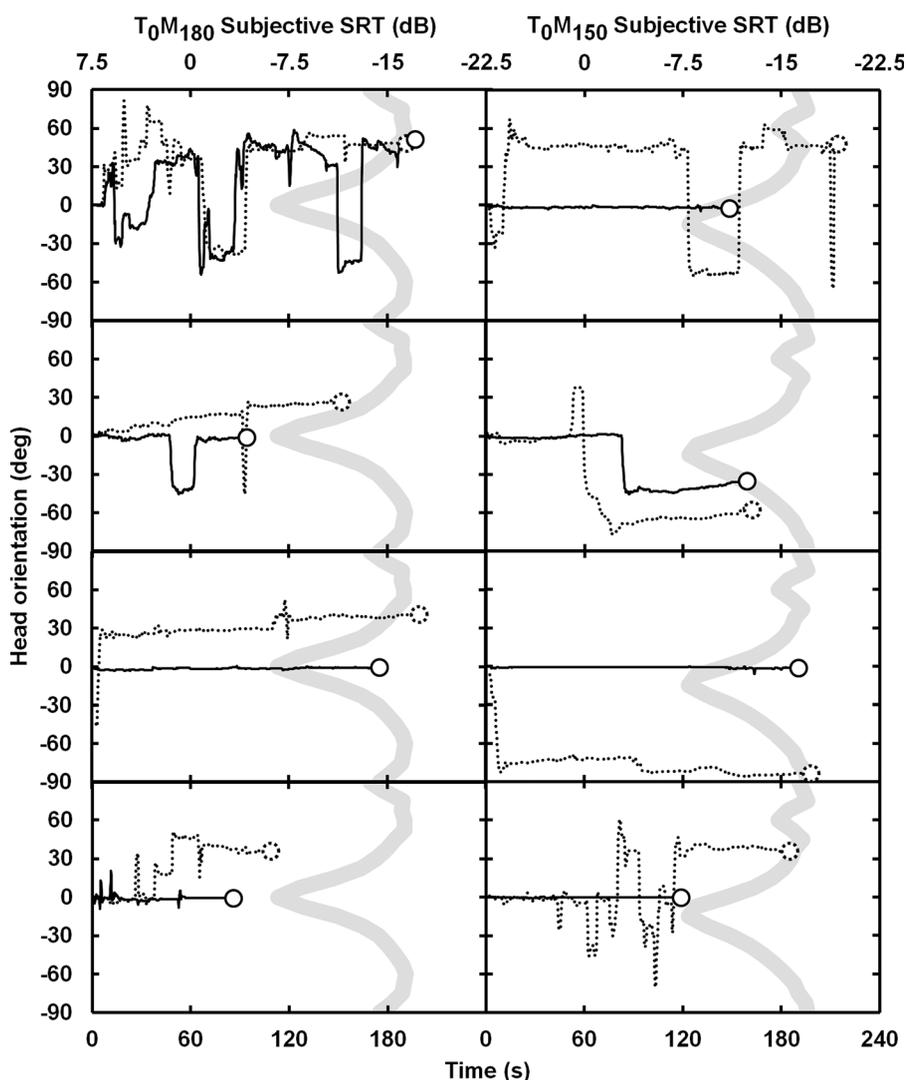


FIG. 3. T_0M_{180} (left panels) and T_0M_{150} (right panels) example head-orientation tracks against absolute clip time (lines) and subjective SRTs (circles) achieved pre- (solid lines) and post-instruction (dashed lines) by four participants. Subjective SRTs are displayed in the context of model predictions (light grey bands) to illustrate how a change in head orientation post-instruction was expected to improve speech intelligibility. Model predictions are inverted and positioned in relation to the x axis by equalization of their mean to the subjective SRT data mean across all runs.

they are much less accurate than the formal SRTs. The head-orientation tracks are displayed in the context of BRIR model predictions (grey bands) so as to illustrate how final head orientations and subjective SRTs might relate to SRM changes. The prediction curve was positioned on the subjective-SRT axis so as to equalize the means of subjective SRTs and predictions across all spatial configurations. The example head-orientation tracks were selected because they illustrate what appear to be different listening strategies. Among all 16 participants, seven did not turn their heads by more than 10° , five turned their heads to the right and four to the left. Of those who did move, some did so erratically and did not necessarily settle at the azimuth providing the largest benefit, even when they might have passed through it at some point in their pursuit. Some moved gradually, thereby gradually improving, but not necessarily reaching the optimum head orientation. Others jumped more or less straight to the most beneficial orientations centered on $\pm 65^\circ$. It was not possible to categorize participants' behaviors, however, since we found no evidence that behaviors did not belong to a continuum. Overall, subjective SRTs loosely followed the predictions of the model.

Example head-orientation tracks in one of the three asymmetrical configurations (T_0M_{150}) are shown in the right-hand side panels of Fig. 3 (solid lines). In the asymmetrical configurations, turning the head to the right was predicted to provide a benefit, whereas turning to the left would be detrimental. Again, some participants did not move and some moved gradually or erratically. Some participants achieved a negative HOB after having passed through a benefit region and some jumped straight to the region of maximum benefit. Again, subjective SRTs loosely followed the model in the three asymmetrical configurations. In 23 out of 48 trials, listeners remained within 10° of facing the speech. Those who moved did not turn significantly more frequently to the right (14) than to the left (11). Listeners' head orientation strategies turned out to be too heterogeneous for us to be able to distinguish between the potential strategies listed in the introduction.

The amount of head movement over each run was computed as the average unsigned head orientation. Head movements appeared to be larger in the T_0M_{180} and $T_0M_{97.5}$ configurations. However, an ANOVA for the amount of head movement across all four spatial configurations resulted in a non-significant effect of configuration [$F(3,42) = 2.55$, $p = 0.069$]. A similar ANOVA for subjective SRTs revealed a significant effect of spatial configuration [$F(3,42) = 10.06$, $p < 0.01$]. The mean subjective SRT was on average 5 dB lower in the asymmetric configurations. Speech-facing SRM changed markedly with masker separation and the HOB contribution to SRM was strongest in the T_0M_{180} configuration. Had listeners reached optimal HOBs, no effect of configuration on subjective SRTs should have been found, because at optimal head orientations the model predicts 10 dB SRM in all four spatial configurations. This further illustrates that listeners were poor at spontaneously reaching optimal HOB.

2. Post-instruction head movements

Only a subset of the listeners (10 of 16) were tested post-instruction. As a result, conditions were not fully

rotated across participants. Post-instruction examples of head orientation tracks are displayed in Fig. 3 (dotted lines). Instruction to explore the benefit of head orientation led all listeners to move away from the speech and had a wide range of effects. Some participants kept a similar strategy to that observed pre-instruction. Others moved their heads much more widely than pre-instruction, sometimes rotating the chair all the way around or turning widely back and forth to either side of the speech, as though to compare the effect of different orientations more quickly. Again, it was not possible to categorize strategies from the range of head tracks observed. An ANOVA operated across all four spatial configurations and comparing pre- and post-instruction head movements (as defined above) revealed a significant increase of head movements with instruction [$F(1,8) = 6.30$, $p = 0.036$] with no effect of configuration.

3. Subjective SRTs

Within each track, the final head orientation was defined as the head orientation averaged over the last 10 s of a head track. The subjective SRT data were very noisy, presumably due to variation in the criteria used by listeners to judge that they had lost track of the meaning of the clips. Nonetheless, combining all trials, a significant correlation was found between the subjective SRTs and SRT predictions for the final head orientations [$r = 0.51$, $t(101) = 5.96$, $p < 0.001$]. This confirms that the more listeners exploited head orientation the more they could understand of the clip. The slope of the regression line was 0.83 (0.14 standard error), again indicating somewhat less benefit of head orientation than predicted.

When comparing pre- and post-instruction subjective SRTs, the mean improvement across 36 pairs of trials was 3.3 dB [$t(35) = 3.91$, $p < 0.001$], suggesting that instruction to explore the potential benefit of head orientation away from the speaker led to speech intelligibility improvement. An important caveat on this outcome is presented in the discussion. When comparing pre- and post-instruction SRM predictions for final head orientations, the mean predicted improvement was 2.5 dB [$t(35) = 3.92$, $p < 0.001$].

IV. DISCUSSION

Predictions of the Jelfs *et al.* (2011) model were compared with SRT measurements for a variety of head orientations with respect to the target speech. An excellent match between SRT data and model predictions was found. Both data and predictions indicated that large HOBs are available in simple listening situations with a single interfering source and modest reverberation. Despite the use of a sound-deadened room, predictions of HOB made from the BRIRs were smaller than those obtained using anechoic HRIRs from KEMAR (Gardner and Martin, 1995). Maximum benefit was predicted to drop from 14 to 9 dB. We concluded that even a modest level of reverberation ($RT60 = 60$ ms) must be responsible for this reduction.

When attending to a diminishing speech level in a fixed noise level and across four different spatial configurations, listeners were found to make use of head orientation in about

half of the undirected trials. Across all four configurations listening behaviors were extremely varied and could not be cleanly categorized. However, it is clear that only half of the listeners who moved made use of an effective listening strategy. Few listeners seemed to know how to make use of sound localization effectively to optimize their strategy. Head movements significantly increased post-instruction, corresponding with a lowering of the mean subjective SRTs reached.

A. HOB comparison with the findings from Kock (1950)

HOB in one of our configurations (T_0M_{180}) was investigated by Kock (1950). He also acquired measurements of HOB in the T_0M_{270} and T_0M_{90} configurations. When comparing the data from Kock (1950) over three spatial configurations with anechoic model predictions, a significant correlation was found [$r=0.76$, $t(22)=5.49$, $p<0.001$, regression slope = 0.73]. This correlation was much weaker and the regression slope even shallower, than found between BRIR predictions and our SRT data. However, it is worth noting that the lowest SRM minima in Figs. 5–7 of the report from Kock (1950) appear not to be reported for the correct head orientations. For instance, in the T_0M_{180} configuration, it is well known that facing the speech provides better intelligibility than facing the noise because the pinnae partly shelter the ears from sound at the rear. In addition, when adopting a head orientation that places target and interferer on the same cone of confusion, the lowest SRM minimum should be found when the target is in the rear hemifield. However, in each of the three spatial configurations selected by Kock (1950), he reported the reverse. It seems likely that head orientations in all three spatial configurations were mistakenly offset by 180° . Correcting for the suspected offset, the lowest SRM minima moved to the correct head orientations in each of the three configurations and a much improved correlation was found [$r=0.91$, $t(22)=10.05$, $p<0.001$, regression slope = 0.87].

B. Listening behaviors

In the behavioral experiment, some listeners did not move and therefore demonstrated no head-orientation strategy. Consequently, they lost track of the speech earlier than other listeners. Some moved but did not seem to make use of any strategy other than moving their head randomly in search of a better SRM. As a result some performed well, some poorly. The remainder seemed to have a much more developed strategy, which allowed them to move straight to the optimum head orientation(s) without the need for scanning. This may be evidence that those listeners made use of localization of the sound sources and worked out where to position their heads before they moved.

Figure 4 shows for each of the spatial configurations a histogram of the final head orientations adopted before and after instruction. HOB predictions are also displayed to illustrate how HOB may have affected listeners' decision to settle at a particular head orientation. In the T_0M_{180} configuration, where participants moved their heads by more than 10° (pre- and post-instruction), they turned to the right (14) more than twice as often as to the left (6) while they should experience equal benefit either way. In the T_0M_{150} configuration where turning to the right is beneficial and turning to the left detrimental, with maximum SRM slope about facing the speech (0.3 dB°), participants turned to the right (12) twice as often as to the left (6). In the other two asymmetric configurations the numbers turning to the right (20) and to the left (18) were almost equal despite the fact that turning to the right was more beneficial, as demonstrated by the SRT data. Overall, when HOB is symmetrical, there seemed to be a bias toward presenting the left ear to the target and when HOB is asymmetrical, there was no clear bias anymore. Over the three asymmetrical configurations, the data suggest listeners' strategy may have been driven by SRM slope only where the SRM slope was highest (Figs. 2 and 4 at T_0M_{150}), i.e., when changes in SRM with head movements could be perceived most sensitively. The SRM slope may not have been sufficiently marked in the remaining asymmetrical configurations for inexperienced listeners

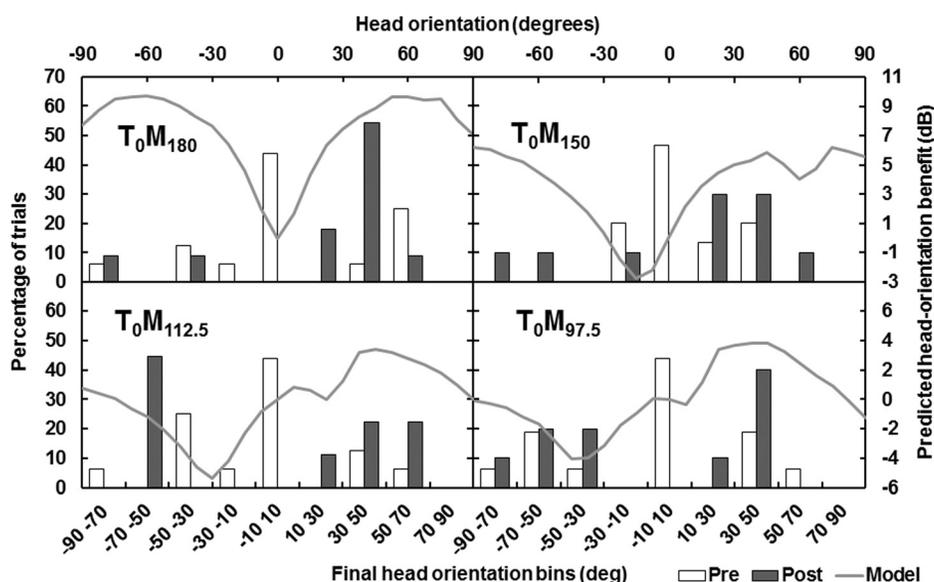


FIG. 4. Histograms of final head orientations adopted in each of the four spatial configurations (T_0M_{180} , T_0M_{150} , $T_0M_{112.5}$, and $T_0M_{97.5}$) pre- (white columns) and post-instruction (grey columns). Model predictions of HOB are also displayed (light grey lines) to assist the reader in judging whether head orientation was driven by changes in SRM with head turns.

to adopt an effective strategy. Hence, our data present only weak evidence that sensing the SRM slope whilst scanning for intelligibility improvement can motivate participants to rotate their head the correct way.

The model shows that a correct strategy is typically to point one's head toward the mid-point between the target speech and the masking noise. Therefore in all three asymmetric configurations where the masker is presented in the listener's right hemifield, turning to the right leads to a positive HOB whilst turning to the left leads to a negative HOB. Since masker positions were not counterbalanced across participants by testing in mirror-image configurations, it is not possible to distinguish between a natural response to the asymmetry of cues and a general bias toward turning to the right. However, upon quizzing the participants with respect to their choice of left or right head turns, more than half of listeners who turned their heads to the left, leading to a negative SRM, indicated that they felt that pointing their head away from the noise would help while in fact by doing so their speech intelligibility was worsened. Some of them, even when given a second chance to explore their HOB in the directed paradigm, persisted with turning away from the noise (see Fig. 4). Presumably, they had achieved some degree of localization of the noise source but had failed to exploit that information correctly. It appears from this that a lot can be done to help people improve their head orientation strategy when attending to speech in noise.

One might argue that the propensity to turn one way rather than the other could be affected by the right-ear listening advantage (e.g., Davidson and Hugdahl, 1996; Kimura, 1961; Springer, 1971). In a dichotic presentation of competing verbal stimuli, recall preference and reaction times have been shown to favor stimuli presented to the right ear as opposed to the left. When free to choose which stimulus to recall, preferential responding is found for stimuli presented to the right ear. However, there is little evidence of a right-ear advantage for speech-in-noise recognition tasks (e.g., Dirks and Wilson, 1969). Therefore, although a right-ear preference could potentially lead listeners to initially turn their heads to the left, so as to bring their right ear closer to the target speech, it would not affect speech-in-noise intelligibility changes with head orientation. Moreover, the final head orientations adopted pre- or post-instruction did not indicate a bias toward turning to the left.

Post-instruction, the amount of head movement increased. Caution should be exercised in directly linking this head movement increase with the post-instruction improvement in subjective SRT. The speech material used pre-instruction was simply repeated in the post-instruction phase of the experiment. This will have rendered the second presentation somewhat more intelligible in noise. Therefore, the 3.3 dB subjective-SRT improvement post instruction may not be fully attributed to listeners adopting more effective head orientations. Indeed, the SRT improvement predicted from final head orientations averaged only 2.5 dB, and this may be a more reasonable estimate of the benefit expected from simply suggesting a change in head orientation.

Brimijoin *et al.* (2012) found that asymmetrically hearing-impaired listeners favored maximizing speech level

over SNR at their better ear. Unfortunately, our choice of spatial configurations and lack of statistical power (too few trials) did not allow us to establish whether our listeners favored the same strategy. The original anechoic predictions were somewhat misleading as BRIR predictions show that for each configuration, the head orientation leading to maximum SRM is close to 60°, also leading to maximum speech level. Even when considering listeners who turned their heads to the left, the range of head orientations reached is such that one cannot conclude that they were aiming for the second maximum speech level orientation of -60°.

In 44% of undirected trials, listeners did not move. In a third to half of the asymmetric trials, the listeners who moved turned the wrong way. Young normally hearing listeners were therefore poor at spontaneously making effective use of the cues available to them. This finding may not be entirely surprising since young normally hearing listeners are the part of the population that least need to make use of head orientation to understand speech in most social settings. Only in severely noisy circumstances such as a loud social or industrial setting might they have, possibly even without realizing it, made use of head orientation. Older normally hearing listeners, whose SRM is known to be reduced (Dubno *et al.*, 1984; Glyde *et al.*, 2011; Helfer *et al.*, 2010; Marrone *et al.*, 2008), would be expected to have encountered more challenging speech-in-noise situations and hence to have developed head-orientation strategies over time. One might therefore expect them to make more spontaneous use of head orientation than younger listeners. Hearing impaired listeners in most noisy situations experience great difficulties. They would therefore be expected to more readily and effectively make use of head-orientation strategies. This remains to be confirmed.

C. Analysis of the BRIRs

We set out to manipulate the BRIRs so as to better understand the predicted effect of reverberation. First (floor), second (ceiling) and third (opposite wall) reflections were individually identifiable in all BRIRs and their timing clearly matched the room's dimensions. It was therefore easy to crop the BRIRs down to the direct sound alone, thereby mimicking anechoic HRIRs. Figure 5 compares predictions for the B&K manikin, rotated with and without room reverberation. A good match was found between MIT HRIR predictions and our anechoic condition (RMS error = 0.5 dB) for each spatial configuration.

Close analysis of the BRIRs showed that the largest reflections were the first reflections from the floor and the ceiling (typically 10–15 dB below the direct sound). The third largest reflection was from the wall opposite to the sound source, but since the walls were strongly sound-treated, this reflection was much weaker than the first two (typically 20–24 dB below direct sound). Cropping of BRIRs so as to include the first three reflections led to predictions within <0.3 dB of uncropped-BRIR predictions. The first three reflections are therefore responsible for most of the effect of reverberation on the HOB and secondary reflections have a negligible impact. Including only the first two reflections rendered the effect of reverberation negligible up to

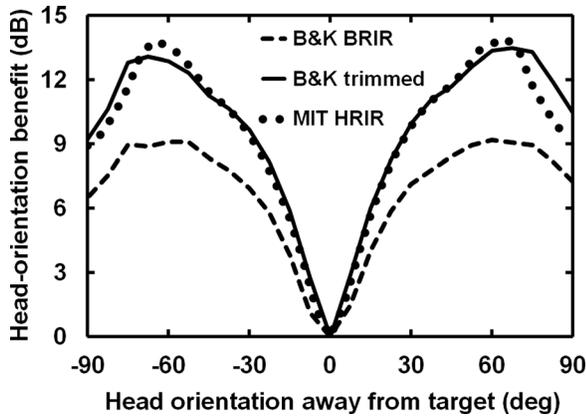


FIG. 5. T_0M_{180} MIT HRIR predictions (dotted line), B&K BRIR predictions (dashed line) and predictions from B&K BRIRs trimmed to direct sound (solid line).

30° head orientation. The third reflection, although much weaker than the first two, is therefore responsible for the majority of the effect for head orientations below 30° . Indeed, as the first two reflections come from the same azimuths as the direct sound (but different elevations), they should not affect interaural time delays (ITDs) and interaural level differences (ILDs) in a manner detrimental to SRM because they remain coherent with the direct sound. The lateral reflections, although weaker, directly affect ITDs and ILDs. The above illustrates how large an effect on SRM/HOB very modest reverberation can have and how important it is that publications that report measures of SRM characterize the impact of reverberation in the test environment, even when reverberation times are as low as that measured in our sound-deadened room (60 ms).

D. Ecological relevance of our findings for an optimum listening strategy

Since this study is the first report of NH HOB and spontaneous listening behavior, it was restricted to the simplest, single-interferer situation. The validation of the Jelfs *et al.* (2011) model for HOB enables us to predict the optimum listening strategy for this simple case and in modest reverberation conditions. First, the listener should position themselves, where possible, such that they are as close to the speech and as far from the interferer as possible and such that the interferer lies in the opposite hemifield from the speech. Once the optimum position is adopted, the listener can then adjust HOB to reach their optimum SRM (10 dB in our modest reverberation condition). This is achieved by rotating the head such that the better ear moves closer to the speech whilst it remains in the shadow of the head with respect to the interferer.

This preliminary study purposely excluded mid-to-high reverberation, fluctuating maskers (including interfering voices), multiple and/or diffuse maskers and increased distance to the target speaker. Therefore, natural extensions to it would consist of introducing the above parameters, one by one or in combination, so as to assess HOB in more ecologically relevant situations. Previous studies of NH speech-facing SRM illustrate how SRM is reduced by azimuthally

distributed interferers, and particularly so when those are distributed over both hemifields (e.g., Peissig and Kollmeier, 1997; Hawley *et al.*, 2004). A free head might enable a listener to move most, if not all interferers to one hemifield so as to place the better ear in their head shadow, which may provide significant HOB.

Although it is known that increasing reverberation reduces both better-ear and binaural-unmasking contributions to SRM, Culling *et al.* (2012) went some way toward illustrating how robust better-ear listening is with reverberation. In the case where reverberation turns multiple interferers into what approaches a diffuse noise, it is arguable that searching for a head orientation that maximizes the target speech level may be the most effective listening strategy. In the most adverse of real-life situations (where speech is nearby and speech and noise are not collocated), such a strategy should always provide a HOB. This may go some way toward explaining why the asymmetrically hearing-impaired listeners from Brimjoin *et al.* (2012) favored maximizing signal level over SNR.

E. Relevance of model and NH findings to HI listeners

Not only do HI listeners suffer elevated intelligibility thresholds, their SRM (and therefore HOB) is also reduced. For them, even a reduced HOB could be critical to their being conversationally included in noisy social settings. The Jelfs *et al.* (2011) model, coupled with an extension of the present study to HI populations, could potentially help define simple guidelines for hearing aid and cochlear implant users to exploit their devices to the full by optimum positioning in a room followed with optimum head orientation. Culling *et al.* (2012) demonstrated how well the Jelfs *et al.* (2011) model predicts the maximum SRM available to unilateral CI users, for whom there is no binaural unmasking. The condition with symmetrical azimuthal separation of target and interferer highlights how CI users could reap the maximum head-shadow benefit predicted by the model at a 60° head orientation in the T_0M_{120} configuration. This was the first demonstration of how well the model could lend itself to predicting SRM for CI users.

Studies of SRM typically limit themselves to fixed-head situations. Not only are free-head situations more ecologically relevant, extending the present study to audio-visual situations would be particularly relevant to HI listeners since they rely most heavily on lip-reading. Maintaining comfortable lip-reading whilst rotating the head away from the speaker implies a sidelong look at the speaker, in which case the range of comfortable head orientations (and compensating gaze angles) may restrict the amount of HOB reachable without compromising lip-reading. Luckily, the model predicts that in most situations where the interferer lies in the rear hemifield, a modest 30° head orientation, presumably compatible with lip-reading, may provide the bulk if not all of the HOB available. It remains to be demonstrated that HOB and lip-reading benefit can be combined.

V. CONCLUSION

The presented study has shown that there are substantial benefits to speech understanding in noise available from orienting the head appropriately (HOB). It further validates the

prediction of the Jelfs *et al.* (2011) model of SRM for HOB. In moderately reverberant conditions, objective measures of SRT showed that HOB could reach up to 8 dB for young normally hearing listeners and that the model could readily predict such benefit within typically 1 dB regardless of masker separation. In a free-head paradigm, listeners did not make optimal use of cues available to them in order to reap HOB. In 44% of trials listeners did not move their head and kept facing the speech. Of those who moved, a few seemed to jump straight to near-optimum orientations, while others moved gradually, erratically and even inappropriately. Repeating the task after instruction, listeners adopted head orientations that provided an SRM improvement overall. This indicated that training on how to optimize one's head orientation strategy might be beneficial, practical, and fast. Extending this study to HI listeners may be particularly important, since they would welcome any strategy that might help them better understand speech in noisy social settings.

ACKNOWLEDGMENTS

We would like to thank our participants for their time and Action on Hearing Loss for funding this research.

- Brimijoin, W., Mcshefferty, D., and Akeroyd, M. (2012). "Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task," *Hear. Res.* **283**, 162–168.
- Bronkhorst, A., and Plomp, R. (1988). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *J. Acoust. Soc. Am.* **83**, 1508–1516.
- Bronkhorst, A., and Plomp, R. (1990). "A clinical test for the assessment of binaural speech perception in noise," *Int. J. Audiol.* **29**, 275–285.
- Bronkhorst, A., and Plomp, R. (1992). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," *J. Acoust. Soc. Am.* **92**, 3132–3139.
- Cherry, E. (1953). "Some experiments on the recognition of speech, with one and with two ears," *J. Acoust. Soc. Am.* **25**, 975–979.
- Culling, J., Jelfs, S., Talbert, A., Grange, J., and Backhouse, S. (2012). "The benefit of bilateral versus unilateral cochlear implantation to speech intelligibility in noise," *Ear Hear.* **33**, 673–682.
- Davidson, R., and Huggdahl, K. (1996). "Baseline asymmetries in brain electrical activity predict dichotic listening performance," *Neuropsychology* **10**, 241–246.
- Dirks, D., and Wilson, R. (1969). "The effect of spatially separated sound sources on speech intelligibility," *J. Speech Hear. Res.* **12**, 5–38.
- Dubno, J., Dirks, D., and Morgan, D. (1984). "Effects of age and mild hearing loss on speech recognition in noise," *J. Acoust. Soc. Am.* **76**, 87–96.
- Duda, R., and Martens, W. (1998). "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.* **104**, 3048–3058.
- Gardner, W., and Martin, K. (1995). "HRTF measurements of a KEMAR," *J. Acoust. Soc. Am.* **97**, 3907–3908.
- Glyde, H., Hickson, L., Cameron, S., and Dillon, H. (2011). "Problems hearing in noise in older adults: A review of spatial processing disorder," *Trends Amplif.* **15**, 116–126.
- Hawley, M., Litovsky, R., and Culling, J. (2004). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," *J. Acoust. Soc. Am.* **115**, 833–843.
- Helfer, K., Chevalier, J., and Freyman, R. (2010). "Aging, spatial cues, and single-versus dual-task performance in competing speech perception," *J. Acoust. Soc. Am.* **128**, 3625–3633.
- Jelfs, S., Culling, J., and Lavandier, M. (2011). "Revision and validation of a binaural model for speech intelligibility in noise," *Hear. Res.* **275**, 96–104.
- Kendon, A. (1967). "Some functions of gaze-direction in social interaction," *Acta Psychol.* **26**, 22–63.
- Kimura, D. (1961). "Cerebral dominance and the perception of verbal stimuli," *Can. J. Psychol.* **15**, 166–171.
- Kock, W. (1950). "Binaural localization and masking," *J. Acoust. Soc. Am.* **22**, 801–804.
- Koehnke, J., and Besing, J. (1996). "A procedure for testing speech intelligibility in a virtual listening environment," *Ear Hear.* **17**, 211–217.
- Lavandier, M., and Culling, J. (2010). "Prediction of binaural speech intelligibility against noise in rooms," *J. Acoust. Soc. Am.* **127**, 387–399.
- Marrone, N., Mason, C., and Kidd, G. (2008). "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," *J. Acoust. Soc. Am.* **124**, 3064–3075.
- Miller, G. (1947). "The masking of speech," *Psychol. Bull.* **44**, 105–129.
- Moore, B., and Glasberg, B. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753.
- Müller, C. (1992). "Perceptive analysis and development of a rhyme test method for speech audiometry," Ph.D. thesis, Gottingen University, Gottingen, Germany.
- Peissig, J., and Kollmeier, B. (1997). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," *J. Acoust. Soc. Am.* **101**, 1660–1670.
- Plomp, R. (1976). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," *Acustica* **34**, 200–211.
- Plomp, R. (1986). "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired," *J. Speech Hear. Res.* **29**, 146–154.
- Plomp, R., and Mimpen, A. (1979). "Improving the reliability of testing the speech reception threshold for sentences," *Int. J. Audiol.* **18**, 43–52.
- Plomp, R., and Mimpen, A. (1981). "Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences," *Acustica* **48**, 326–328.
- Rothausen, E., Chapman, W., and Guttman, N. (1969). "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoustics* **17**, 225–246.
- Schorr, E., Fox, N., van Wassenhove, V., and Knudsen, E. (2005). "Auditory-visual fusion in speech perception in children with cochlear implants," *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18 748–18 750.
- Springer, S. P. (1971). "Ear asymmetry in a dichotic detection task," *Percept. Psychophys.* **10**, 239–241.
- Sumbly, W., and Pollack, I. (1954). "Visual contribution to speech intelligibility in noise," *J. Acoust. Soc. Am.* **26**, 212–215.