“Turn an ear to hear”: How hearing-impaired listeners can exploit head orientation to enhance their speech intelligibility in noisy social settings

JACQUES A. GRANGE[1,*], JOHN F. CULLING[1], BARRY BARDSLEY[1],
LAURA I. MACKINNEY[1], SARAH E. HUGHES[2], AND STEVEN S. BACKHOUSE[2]

[1] *School of Psychology, Cardiff University, Cardiff, UK*

[2] *South Wales Cochlear Implant Programme, Princess of Wales Hospital, Bridgend, UK*

Head orientation enhances the spatial release from masking. Here, with their head free, listeners attended to speech at a gradually diminishing signal-to-noise ratio (SNR) and with the noise source azimuthally separated from the speech source by 180 or 90°. Young normal-hearing listeners spontaneously turned an ear towards the speech source to improve speech intelligibility in 64% of audio-only trials, but a visible talker’s face and/or cochlear implant use significantly reduced this head-turn behaviour. Instructed to explore the potential benefit of head turns, all listener groups made more head movements and followed the speech to lower SNRs. Unilateral CI users improved the most. In a virtual restaurant simulation with 9 interfering noises/voices, hearing-impaired listeners and simulated bilateral CI users typically obtained a 1-3 dB head-orientation benefit from a 30° head turn away from the talker. In this diffuse interference, the effect is due to improved target level rather than reduced noise at the better ear. Surveys of UK CI users, CI clinicians and internet-based communication advice, showed that most advice was to face the talker head on. CI users would benefit from guidelines that recommend looking sidelong to present their better hearing implanted ear towards the talker.

INTRODUCTION

Spatial release from masking (SRM) improves intelligibility through spatial separation of target speech and interfering sources. Typically, listeners face the speech head on. It was assumed by researchers and professionals that facing the speech was a more natural attitude (Bronkhorst and Plomp, 1990). However, Kock (1950) found a large benefit of orienting the head away, a benefit also predicted by the Jelfs *et al.* (2011) model of SRM. Grange and Culling (2016a) demonstrated that young normal-hearing (NH) listeners could obtain as much as 8 dB improvement in speech reception threshold (SRT) in a sound-treated room. Most of this head-orientation benefit (HOB) was obtained with a modest 30° head turn. Grange and Culling (2016b) showed that a significant HOB was also obtained by CI users alike, with a 30° turn and 180° or

---

*Corresponding author: grangeja@cardiff.ac.uk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017. ISBN: 978-87-990013-6-1.

90° source separation. Unilateral CI users obtained the same HOB (up to 4.5 dB) as age-matched NH controls. Bilateral CI users obtained less but still significant HOB (up to 2 dB). Testing listeners in audio-visual modality (AV) in addition to audio-only (A), Grange and Culling (2016b) confirmed that a 30° head turn had no detrimental impact on the listeners' lip-reading ability. Therefore, for CI users, the benefits of head orientation and lip-reading could be combined to improve SRT by up to 9 dB.

Grange and Culling (2016b) also tested whether HOB would occur in a typical noisy and reverberant setting. A realistic restaurant simulation was created using binaural room impulse responses from a B&K head-and-torso simulator in a real restaurant. The manikin was sat at 6 different tables with its head in 3 different orientations. Small loudspeakers represented a talker sat across the table and 9 concurrent interferers throughout the restaurant. SRT measurements over headphones showed that NH listeners benefited from a 30° head orientation and, on average, gained ~1.5 dB at the predicted best head orientation. Culling (2016) showed that such a benefit was not due to the acoustic shadow of the head (how head-shadow is most often understood) but instead mostly due to an amplification of the target level at the better ear.

The present study extends those of Grange and Culling (2016a,b) in two ways: First, we investigated head-orientation behaviour of CI users when the head is free, and second, we measured the HOB of hearing-impaired listeners and simulated CI users in the restaurant simulation. Experiment 1 tested (1) whether listeners spontaneously turn their heads when attending to a target talker in noise and (2) whether a simple instruction to explore their HOB leads to better performance. Grange and Culling (2016a) had already showed that in 56% of audio-only trials, listeners spontaneously turned their heads but did not adopt ideal orientations. This may in part be explained by Brimijoin *et al.*'s (2012) finding that asymmetrically impaired listeners tended to optimize target level rather that SNR at their better ear. Experiment 2 tested HOB in the simulated restaurant, for HI listeners and simulated CI users.

EXPERIMENT 1: HEAD-ORIENTATION BEHAVIOURAL TASK

Participants

The same participants as in Grange and Culling (2016b, Expt. 1) were tested according to the rules of our institutional Ethics Committee: 12 young NH listeners, 16 CI users (8 unilateral and 8 bilateral) and 10 NH listeners age-matched to the CI users.

Spatial configurations

The free-head task was run in both the collocated (T_0M_0) and the separated (T_0M_{90} and T_0M_{180}) spatial configurations, so that subtracting a separated-configuration SRT from the collocated-configuration SRT (within a presentation modality) would lead to a measure of SRM. Figure 1 illustrates the Jelfs *et al.* SRM model predictions as a function of head orientation for the separated spatial configurations.

Stimuli

Passages from the *The Wonderful Wizard of Oz* were audio-visually recorded. Each 3-4 s segment of the audio stream was normalised for RMS power. Gaps in speech exceeding 100 ms were excluded from the RMS calculation. Masking noise was filtered to match the long-term spectrum of the voice.

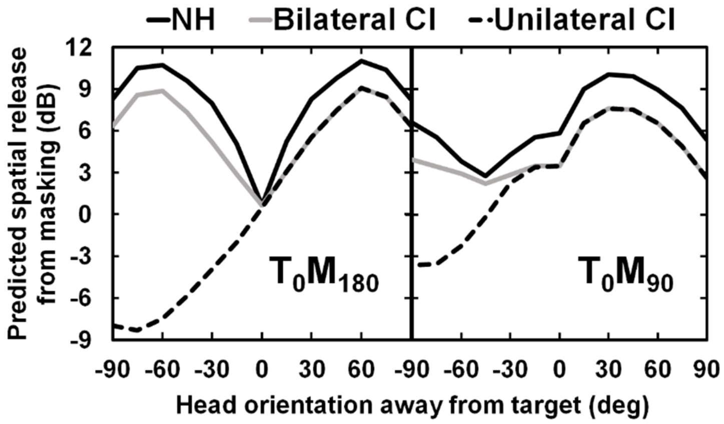

Fig. 1: Predicted spatial release from masking for all listener groups as a function of head orientation and for maskers at 180° or 90° azimuth.

Audio and AV protocol

Listeners were presented with a set of six 6-minute long clips, starting at SNRs of 6 dB for NH and 16 dB for CI users. SNRs diminished at 6 dB/min., so that the SNR would reach the listener's 50% intelligibility point in the collocated condition about two minutes into a clip and no listener could follow a clip all the way to its end. As in Grange and Culling (2016a), listeners were instructed to listen "normally as in a social situation" but "keep your back against the chair's back rest and keep your arms resting on your lap". Listeners were told they would be quizzed on the last 3-5 words they felt they correctly understood in sequence. Presentation stopped when listeners flagged that they had lost track of the thread of the clip, and they recalled the last 3-5 words. The listeners were not told where the target speech would come from, but they spontaneously faced the video monitor at the start of each trial. Next, the listener was informed that head orientation might be beneficial and repeated the free-head test, making use of the same material. The rest of the instructions remained the same as for the first test.

Results

Overhead video recordings were post-processed using MATLAB to recover head orientation. Over two passes, an operator tracked with the mouse 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. The recalled words were located in the clip's transcripts to estimate SRT: the SNR at which listeners lost track of the clips.

Analysis of the variance of the amounts of head movements (mean, unsigned head orientation) as a function of group, presentation modality, spatial configuration and instruction revealed significant increase of head movements after instruction [$F(1,34) = 179.2$, $p < 0.001$] and inhibition of head movements by AV presentation [$F(1,34) = 91.6$, $p < 0.001$]. AV presentation had a greater inhibiting effect on CI users than NH listeners [$F(3,34) = 221.2$, $p < 0.05$]. Instruction reduced the inhibiting effect of the AV modality [$F(1,34) = 7.1$, $p < 0.05$], particularly for CI users [$F(3,34) = 3.8$, $p < 0.05$]. Differences between spatial configurations were also removed by instruction [$F(2,68) = 3.5$, $p < 0.05$].

Where NH listeners employed head movement, most scanned for intelligibility improvements but few went straight to the predicted most beneficial head orientations. CI users made more conservative head turns and never went straight to the predicted most beneficial head orientations, perhaps because of their poorer sound localisation ability (Kerber and Seeber, 2012). Of NH listeners who scanned for improvement, most settled at sub-optimal orientations, even after passing through optimal orientations. Unilateral CI users mostly turned the correct way after instruction. However, for T_0M_{90}, more than half of age-matched NH listeners and bilateral CI users turned away from the noise, as though they had tried to get away from it, when the optimal strategy was to point their head between speech and noise directions.

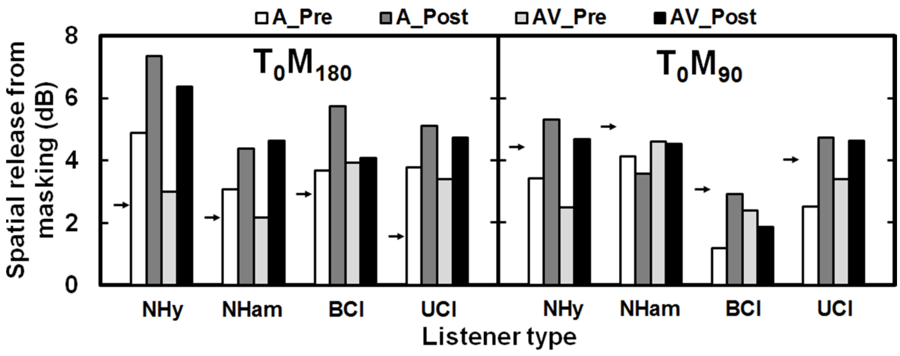

Fig. 2: SRM reached at final head orientation by each group [young NH (NH_y) and age-matched NH (NH_{am}) listeners; bilateral CI (BCI) and unilateral CI (UCI) users], in each spatial configuration for audio-only (A) and audio-visual (AV) presentation modalities, pre-instruction and post-instruction. Arrows highlight the speech-facing SRMs from Grange and Culling (2016b).

Figure 2 presents the mean pre-instruction and post-instruction SRM for each listener group and in each spatial configurations. Pre-instruction, listeners performed better at T_0M_{180} than if they had remained facing the speech. This is to be expected since for most listeners HOB could be had from any head turn away from the speech. At T_0M_{90}, however, listeners spontaneously performed worse than if they had remained still. Overall, young NH and age-matched NH listeners and CI users all improved as a result of instruction [by 1.6, 0.8 dB and 1.2 dB, $F(1,11) = 7.80$, $p < 0.02$; $F(1,9) = 11.05$, $p < 0.01$ and $F(1,15) = 5.27$, $p < 0.05$, respectively]. Significant correlations between subjective SRMs and SRM predictions at final head orientations were found for each listener type [$r = 0.49$, $t(46) = 3.86$, $p < 0.001$; $r = 0.35$, $t(38) = 2.30$, $p < 0.03$; $r = 0.36$, $\mathrm{t}(60) = 2.96$, $p < 0.005$, respectively], indicating that improved head orientations led to SRM improvement. Despite an overall positive effect of instruction, age-matched NH listeners and bilateral CI users did not improve post-instruction at T_0M_{90}.

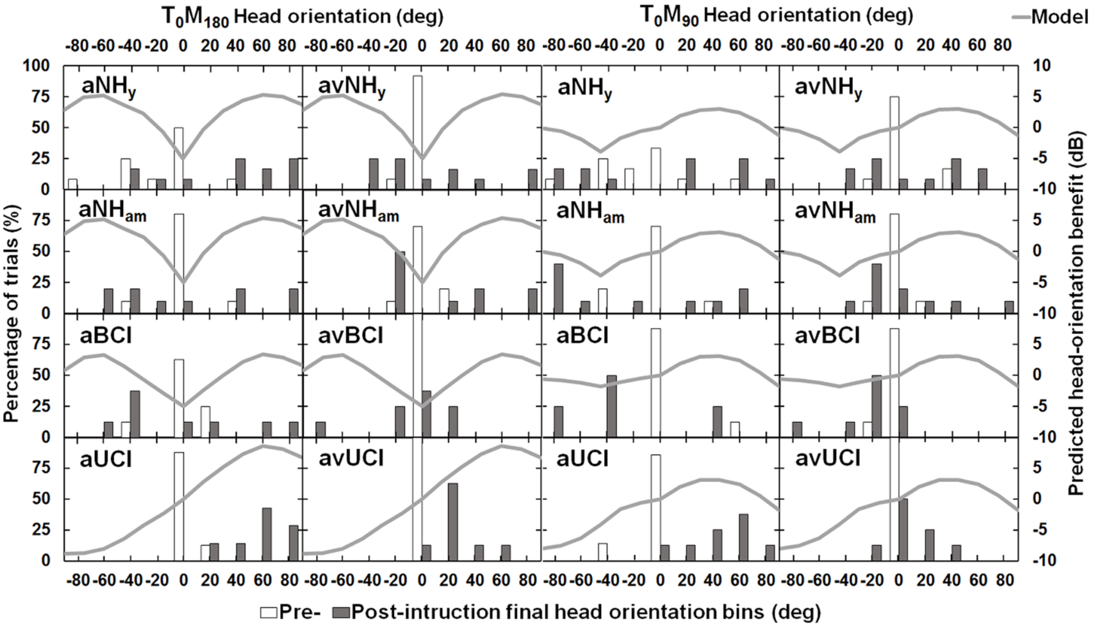

Fig. 3: Histograms of the final head orientations of young NH (NH_y), age-matched NH (NH_{am}), bilateral CI (BCI) and unilateral CI (UCI) listeners for audio-only (a) and audio-visual (av), pre (white bars) and post (dark grey bars) instruction. Predicted SRMs are light grey lines.

Figure 3 shows histograms of final head orientations for each listener group at T_0M_{180} and T_0M_{90}. Model predictions are superimposed to help the reader judge how well listeners discovered optimal HOB pre- and post-instruction. The inhibition of head movements by the presence of visual cues is demonstrated by the tight distribution of final-head-orientations around the speech-facing orientation in the AV modality. At T_0M_{180}, NH and bilateral CI users can get a benefit of turning either way. Unilateral CI users, however, need to turn to present their implanted ear, and it is clear that they all turned the correct way. At T_0M_{90}, all listeners should experience a benefit of pointing their head between speech and noise sources. In only 1 of 16 post-instruction trials, did a unilateral CI user turn the wrong way. In contrast, bilateral listeners settled at detrimental orientations in 12 out of 16 trials, age-matched NH listeners in 11 of 20 trials and young NH listeners in 9 of 24 trials.

EXPERIMENT 2: SIMULATIONS IN A VIRTUAL RESTAURANT

The materials from Grange and Culling's (2016b) second experiment were employed. Listeners and target talkers were sat across the table from each other at each of 6 tables in a virtual restaurant. Interferers came from another 9 tables spread across the restaurant. Interferers were either steady speech-shaped noise or continuous voices. The combination of the 9 interferers produced a spatialized babble or diffuse noise.

Participants

16 young NH adult listeners (mean 21 y.o.) and 14 unaided HI listeners (mean 68 y.o.) with moderate to severe high-frequency loss (40-85 dB HL in at least one ear and increasing from 4 kHz) participated, in accordance with our Ethics Committee rules.

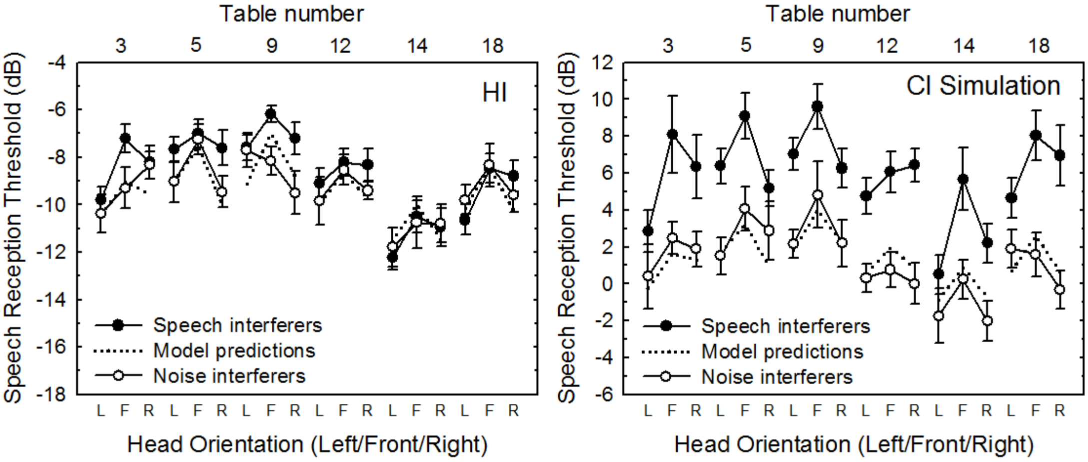

Fig. 4: SRTs obtained for HI listeners (left panel) and simulated CI users in the virtual restaurant, as a function of head orientation (30° to the Left or Right, or Facing the target talker), interferer type (open circles for speech-shaped noise, closed circles for babble) and table. Jelfs *et al.* model predictions (dotted lines), with their mean equalised to mean SRTs in noise, include binaural unmasking for HI listeners, but not for simulated CI users.

Simulation of CI users

The mixed target-interferers signal was passed through SPIRAL, a tone vocoder that incorporates the threshold-elevating effect of current spread (set at 8 dB/oct.) using 80 carrier tones. For details of the vocoder, see Grange *et al.* (2017). NH participants listened to the combined restaurant and CI simulation.

Results with HI listeners and simulated CI users

The left and right panels of Fig. 4 plot SRTs as a function of table number, head orientation (left, front, right) for each interferer type for HI listeners and simulated CI users, respectively. SRTs were ~12 dB higher for simulated CI than HI listeners. The benefit of orienting the head was significant for each listener group [HI, $F(2,26) = 17.4, p < 0.001$; CI, $F(2,36) = 15.1, p < 0.001$]. At the best predicted head orientation and in noise, the magnitude of HOB was 1.2 and 1.7 dB for HI and CI users, respectively. SRTs were significantly higher for speech than noise interferers (by 1 and 5 dB for HI and CI users, respectively). Simulated CI users alone benefitted more from head turns in babble, with 3.2 dB HOB, than in noise.

DISCUSSION

Experiment 1 found that when speech was hard to follow CI users made significantly less spontaneous head turns than NH listeners, particularly with AV present. However, with a simple instruction to explore their HOB, all listener groups could follow the clips to significantly lower SNRs. Our findings suggest that simple training of HI listeners to exploit their HOB could improve their speech understanding in noisy environments. Listeners did not exhibit a clear set of head-orientation behaviours that could be categorised. A study with a much larger sample size would help establish whether behavioural categorisation is justifiable. Such a study could help tailor the design of training programs to each specific listener’s needs.

Experiment 2 tested the robustness of HOB with reverberation and multiple interferers. Regardless of the table position within the restaurant or of the interferer type, a HOB of 1.2-3.2 dB could be obtained. Comparing results to Expt. 2 of Grange and Culling (2016b), SRTs are elevated in HI and simulated CI users. In addition, HI and simulated CI listeners exhibited even higher SRTs when immersed in a spatialized babble than in speech-shaped noise. Qin and Oxenham (2003) concluded from their CI simulations that in order to segregate a target voice from background interferers, both F0 segregation and good frequency resolution were required. While our HI listeners may still, via low frequencies, be able to exploit F0 differences, their limited frequency resolution may explain their higher SRTs with interfering voices. For our simulated CI users, not only is their frequency resolution significantly reduced by CI simulation, but they have no access to the F0 cue. This may explain their greater SRT elevation with voiced interferers. What remains unclear is why they appear to benefit more from head orientation (3.2 dB HOB) than their NH or HI counterparts in a spatialized babble. The data suggest that the modulation or informational masking by babble changes faster with head orientation than energetic masking by noise.

To compare our recommendations with current practice, we surveyed 95 CI users and 31 CI clinicians regarding advice given to CI users about head orientation. 89% of clinicians reported frequently or always advising CI users to directly face the talker; 77% of CI users reported the same. A further survey of communication advice available on the internet found 23 “public information” websites. The majority recommended strategies to reduce background noise and stated, in all but one case, that the listener should face the speaker. The clinicians' responses revealed the two assumptions that mostly influenced their advice: (1) that facing the talker leads to better lip-reading and (2) that one must face the talker to benefit from microphone directionality. However, Grange and Culling (2016b, Fig.7) demonstrated that lip-reading was unaffected by head orientation of 30° away from the talker and that maximum sensitivity of a directional microphone on a BTE hearing prosthesis is in fact shifted to 30° to 50° azimuth by the acoustic diffraction of the head.

Overall, these experiments demonstrate that: (1) HOB in a realistic social setting is robust with moderate-to-severe high-frequency hearing loss or for a simulated CI listener; (2) simulated CI users benefit from HOB as much as NH or mild-to-moderate HI individuals, and (3) CI users and other HI listeners would benefit from the development of listening-strategy training that involves appropriate head orientation.

This research was supported by Action on Hearing Loss and the Oticon Foundation.

REFERENCES

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. doi: 10.1016/j.heares.2011.10.009

Bronkhorst, A., and Plomp, R. (**1990**). "A clinical test for the assessment of binaural speech perception in noise," Int. J. Audiol., 29, 275-285. doi: 10.3109/00206099009072858

Culling, J.F. (**2016**). "Speech intelligibility in virtual restaurants," J. Acoust. Soc. Am., **140**, 2418-2426. doi: 10.1121/1.4964401

Grange, J.A., and Culling, J.F. (**2016a**). "The benefit of head orientation to speech intelligibility in noise," J. Acoust. Soc. Am., **139**, 703-712. doi: 10.1121/1.4941655

Grange, J.A., and Culling, J.F. (**2016b**). "Head orientation benefit to speech intelligibility in noise for cochlear implant users and in realistic listening conditions," J. Acoust. Soc. Am., **140**, 4061-4072. doi: 10.1121/1.4968515

Grange, J.A., Culling, J.F., Harris, N.S., and Bergfeld, S. (**2017**). "Cochlear implant simulator with independent representation of the full spiral ganglion," J. Acoust. Soc. Am., **142**, EL484-EL489.

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. doi: 10.1016/j.heares.2010.12.005

Kerber, S., and Seeber, B.U. (**2012**). "Sound localization in noise by normal-hearing listeners and cochlear implant users," Ear Hearing, **33**, 445-457. doi: 10.1097/aud.0b013e318257607b

Kock, W. (**1950**). "Binaural localization and masking," J. Acoust. Soc. Am., **22**, 801-804. doi: 10.1121/1.1906692

Qin, M., and Oxenham, A. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am., **114**, 446-454. doi: 10.1121/1.1579009