The benefit of cochlear-implant users' head orientation to speech intelligibility in noise

JACQUES A. GRANGE* AND JOHN F. CULLING

Cardiff University, School of Psychology, Cardiff CF103AT, United Kingdom

Speech reception thresholds (SRTs) in noise improve when the speech and noise sources are spatially separated. This spatial release from masking (SRM) is usually investigated in fixed-head situations. We studied free-head situations in audio and audio-visual conditions. We compared normally-hearing and cochlear-implant (CI) users' spontaneous and directed head-orientation strategies when attending to speech in noise with a progressively declining signal-to-noise ratio. SRM-model predictions suggested benefits of head orientation away from the target speech that we hypothesized would motivate head rotation. As signal-to-noise ratio declined, observed head tracks differed greatly between listeners. Audio-visual presentation reduced the amount of head rotation. When directed, listeners made more effective use of head rotation. Audio and audio-visual SRTs were acquired at fixed, 0, and 30 deg head orientations with respect to the target speech. At the most favourable 30-deg head orientation, SRM reached 8 and 6 dB for NH listeners and CI users respectively. Lip-reading yielded improvements of 3 and 5 dB on average across conditions. CI users confirmed that training in optimizing both their position and head orientation with respect to target speaker and noise source position in a social setting was both currently missing and likely valuable.

INTRODUCTION

Bilateral cochlear implantation provides service users with several benefits over unilateral implantation. In addition to sound-source localization being made possible to some extent, Van Hoesel and Tyler (2003) showed that bilateral cochlear-implant users (BCIs) benefit from improved speech intelligibility in noise (SpIN) when speech and noise sources are spatially separated. However most studies to date have considered such spatial release from masking (SRM) in a fixed-head situation (e.g., Van Hoesel and Tyler, 2003; Litovsky *et al.*, 2006; Loizou *et al.*, 2009). Furthermore and with few exceptions, most examined SRM by comparing speech co-located with noise in front of the listener with speech in front and noise azimuthally separated by 90 deg to the left or to the right, configurations known to not make optimum use of the head-shadow effect due to the bright spots located at ±90 deg. Our model of SRM (Jelfs *et al.*, 2011) predicted the spatial configuration providing the maximum benefits of bilateral over unilateral implantation, later confirmed by Culling *et al.* (2012) with normally hearing-listeners (NHs) and cochlear-implant users (CIs). The SRM model could also be used to predict how

*Corresponding author: grangeja@cf.ac.uk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

head orientation away from facing the speech could yield improved SpIN. From the conclusions of Culling *et al.* (2012) one could guide CIs with respect to their optimal seating strategy. Seating options in a restaurant could however be limited and the only degree of freedom left would be head orientation.

In a first experiment we examined in a sound-deadened room whether the model predictions translated to CIs adopting effective spontaneous free-head orientation strategies. We also established whether simple guidance could immediately make a large difference in the lowest speech-to-noise ratio (SNR) they could successfully reach. A baseline was established with NHs. Trials were conducted in audio-only or audio-visual (AV) conditions to measure the impact of lip-reading.

In another two experiments with the same participants, in the same room and spatial configurations, fixed-head speech reception thresholds (SRTs) were measured as well as their improvement by a modest, 30 deg head orientation away from the speech-facing direction. This enabled direct validation of the model predictions without compromising lip-reading, 30 deg being thought as an acceptable gaze angle for lip-reading purposes. In the BCIs case we also measured summation and squelch (Schleich *et al.*, 2004).

MODEL OF SRM

The model of SRM originally introduced by Lavandier and Culling (2010) takes as input speech-shaped noise, which has been convolved with binaural-room-impulse-response recordings to create a reverberant speech-shaped noise target and interferer.

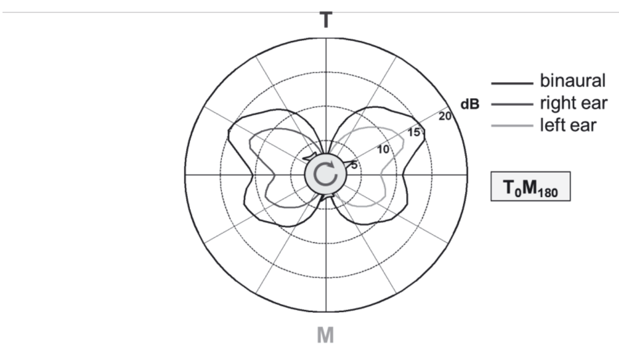

Fig. 1: Predicted head-orientation benefit (anechoic condition) with target in front and masker at 180 deg.

A first path calculates the expected binaural advantage due to binaural unmasking using equalization-cancellation theory to predict the binaural masking level difference. A second path predicts the benefits of better-ear listening or head-shadow effect. Combined, the two paths account for the two cues associated with SRM (Bronkhorst and Plomp, 1988). The two outputs are simply added to generate an SRM prediction. Jelfs *et al.* (2011) refined the model by enhancing its computational efficiency. Culling *et al.* (2012) made use of the model to predict SRM for NHs and CIs with one speech-shaped interfering noise. Figure 1 illustrates the model predictions as a function of head orientation in anechoic conditions. The two inner curves are predictions for left or right ear alone. Because bilateral CI users do not benefit from binaural unmasking, the head-orientation benefit they would experience is the outermost these two curves, i.e., the benefits of better-ear listening. The outer curve is a NH prediction which includes binaural unmasking. Up to 12 and 16 dB benefit is predicted at ±60 degrees head orientation for CIs and NHs respectively.

MATERIALS AND METHODS

Participants

12 NH participants aged from 18 to 22 (age mean: 20) were recruited from the undergraduate Cardiff University population. 9 CIs aged 35 to 72 (age mean: 62) participated, of which 5 were bilateral CIs and 4 were unilateral CIs. CIs were recruited through the UK National CI User Association and used implants from a mix of manufacturers (Cochlear, MedEl, and Advanced Bionics). All CIs had had their last implant fitted 2+ years before being tested.

Laboratory setup

Two sound-deadened rooms were used and acoustically matched. As schematically shown in Fig. 2, 4 × Cambridge Audio Minx loudspeakers were arranged at cardinal positions around a circle of radius 1.3 m centered on the listener's head, driven by a 6-channel Auna solid-state amplifier and an ESI MAYA44+ I/O sound card. A 17-inch screen was positioned below the front speaker, through which the target speech was always presented. A shaving mirror was used to assist listeners in adopting the correct head orientations during the SRT runs. The RT60 of the rooms was derived from impulse measurements to be circa 60 ms. A webcam fitted on the ceiling above the listener's head enabled covert video recording and subsequent extraction of participant head-tracks.

Stimulus presentation and preparation

The speech and noise were presented either directly by Matlab and Playrec or through the VideoLAN player. A set of 320 high-predictability-SPIN-sentence audio-visual clips were recorded to measure the impact of lip-reading. The reading of sections of The Wonderful Wizard of Oz by L. Frank Baum was video-recorded for the free-head task, a material chosen for its predictability. 570 sentences from the Harvard IEEE sentence corpus were used for more precise audio SRT measurements. All audio material was sampled or re-sampled at 44.1 kHz and rms-normalized.

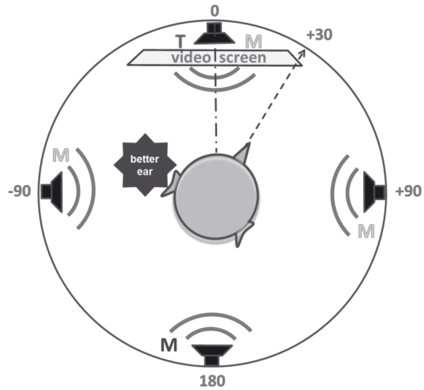

Fig. 2: Highlighted (darker masker or target markers M or T) are the T_0M_{180} spatial configuration and $H_{30}M_{180}$ global configuration.

Spatial configurations

The model predicted that maximum SRM gain could be obtained with target at 0 deg and masker at 180 deg azimuths respectively (the T_0M_{180} configuration). Informed by a prior NH study and given most previous studies tested for SRM with speech in front and masker at ±90 deg, tests were conducted in 16 combinations of spatial configuration, head orientation, and presentation modality: spatial configuration with target and masker at 0 deg vs. target at 0 deg and masker at ±90 deg or 180 deg; head either facing the speech or with the head rotated by ±30 deg; audio or AV. A 90 deg masker separation or 30 deg head turn favoured the listener's best performing ear for speech perception in noise. The model predicted that a favourable 30 deg head turn would provide either the bulk of the attainable audio SRM in the T_0M_{180} configuration or the maximum attainable SRM in the favourable T_0M_{90} configuration. The T_0M_{90} configuration would allow us to correlate our new audio-only data with the Culling *et al.* (2012) study and other prior studies. The T_0M_{180} would maximize benefit of head rotation according to model predictions. T_0M_0 acted as a reference for all other SRT data or as control in the free-head experiment. Spatial configuration and head angle were combined for simplicity into a global configuration code such as the $H_{30}M_{180}$ configuration highlighted in Fig. 2.

RESULTS AND DISCUSSION

Experiment 1: Free-head listening task

The central plot of Fig. 3 is the outcome of the undirected, audio-only condition. SNR at source (proportional to time) is presented radially and head orientation azimuthally. To the left, the change of behaviour reflects the effect of lip-reading; to the right how much or how quickly a listener can learn to make use of head orientation. Arrows highlight the range of optimum orientations. For the sake of brevity only the T_0M_{180} UCIs plots are included here. Each line corresponds to a given participant's head track; a circle is positioned at the SNR/time corresponding to the last 3-5 words correctly understood by the participant, corresponding to a self-reported SRT-50 for their final head angle.

Contrary to 45% of NH listeners who made spontaneous use of head orientation, only 10% of CIs appeared to turn their heads (undirected, audio-only) and the presence of the visual cues totally eradicated head turns. Once directed, most CIs achieved a very significant improvement, most reaching the optimum SRM orientation(s). No CI turned their heads when the speaker's face was visible and all reached between 5 and 10 dB more intelligibility when directed. 90% of CIs reached the very best performance by combining head orientation and lip-reading in the AV directed task. Most reached 20-25 deg away from the speech direction.

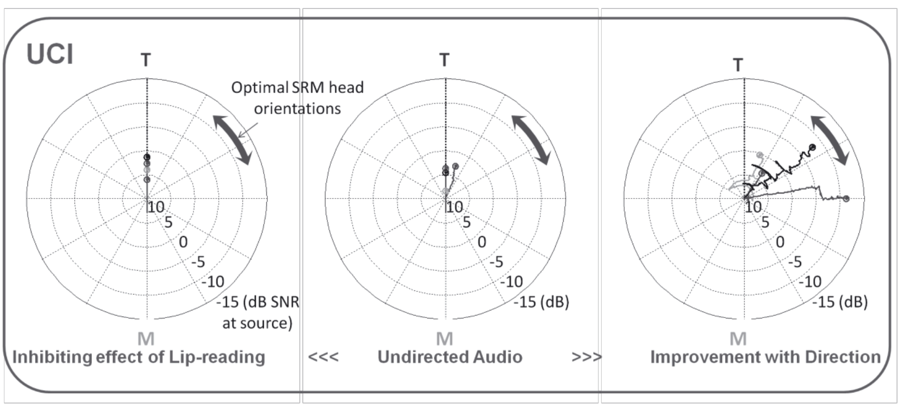

Fig. 3: UCI example of head orientation tracks.

Experiment 2: Gain of lip-reading

In each spatial configuration and making use of the SPIN sentence material, the SRTs acquired in audio-only were subtracted from SRTs acquired in the AV mode. Figure 4 shows the benefit that lip-reading provided over and above SRM benefits. NH's and CI's lip-reading gains were respectively found to be 3 and 5 dB on average across conditions. Moreover and most importantly, lip-reading was confirmed to be beneficial in all configurations and a 30 deg head turn did not significantly and adversely affect lip-reading.

Experiment 3: SRM, summation, and squelch findings

Figure 5 presents on the right hand side the SRM results obtained for NHs and BCIs. Dotted and dashed lines are the model predictions for each group, continuous lines and symbols are means across participants. The error bars reflect the standard error of the means.

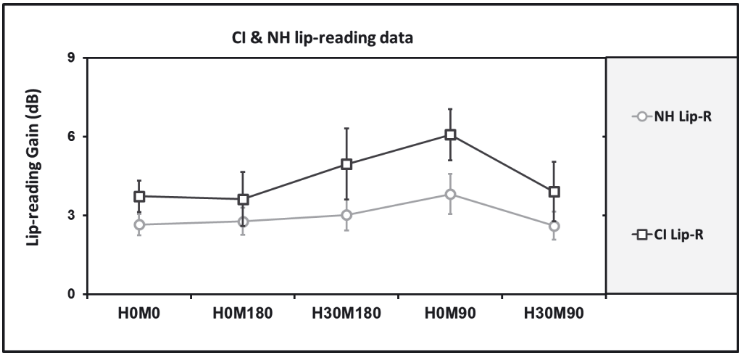

Fig. 4: NHs and CIs lip-reading gain.

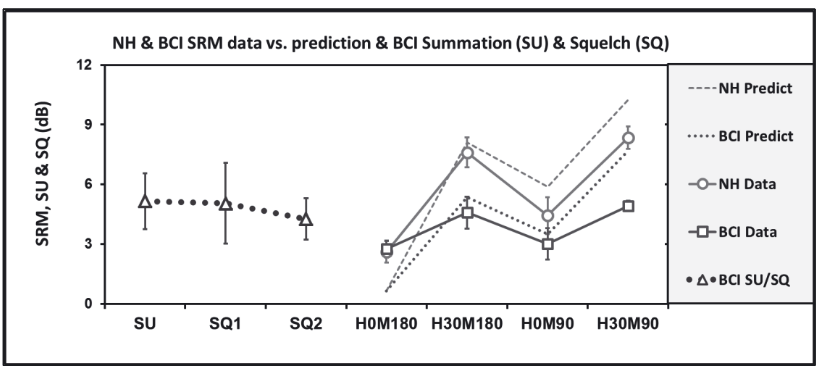

Fig. 5: NHs SRM; BCIs SRM, summation and squelch.

The predicted benefit of head rotation from H_0M_{180} to $H_{30}M_{180}$ is noticeably larger for NHs (7.4 dB) than for BCIs (4.3 dB) and so solely due to binaural unmasking. NHs and BCIs all benefitted from a 30-deg head turn in both spatial configurations by up to 4.5 dB and 1.9 dB for NHs and BCIs, respectively. The lower than predicted CI 30-deg head-turn gain may be attributable to the variety of microphone positions in the various implants used by our participants; the further away from the KEMAR manikin's microphone position in the ear canal, the larger the effect. The discrepancy is indeed expected to be attributable to variations in head-shadow effect. On the left hand side of Fig. 5, summation and squelch data are displayed. These are the SRT improvement between the best ear/implant performance (second implant disabled) and performance with both cochlear implants on. The summation is calculated from H_0M_0 SRT data whereas the squelch is calculated from the H_0M_{90} or $H_{30}M_{180}$ as per Schleich *et al.* (2004). Both summation (5.2 dB) and squelch (4.2 and 5 dB) mean values are far superior to the 1-2 dB reported elsewhere (e.g., Schleich *et al.*, 2004). The two BCIs showing largest summation and squelch reported large differences in the character of the sound perceived from each implant, suggesting a spectral summation effect rather than any binaural unmasking.

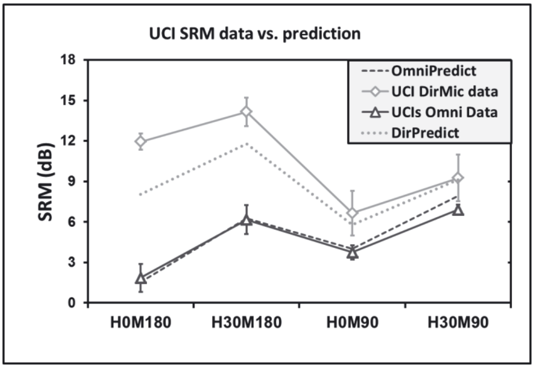

Fig. 6: SRM data vs. predictions for UCIs, omni and directional microphones.

Figure 6 illustrates the quality of match between predictions and data for the omni-directional microphone UCIs mean SRMs. All means were indeed within 1 dB of predictions. 4.5 dB was gained from a favourable 30 deg head turn.

One UCI used a directional microphone setting, which had the effect of boosting their SRM by over 10 dB in the T_0M_{180} configuration. The relatively small discrepancy (3 dB) between data and omnidirectional predictions with a 90-deg masker separation however shows that these directional settings are not so influential for sound towards the front. New directional predictions were generated by adding the computed difference between directional and omnidirectional anechoic predictions (Cochlear HRIRs) to the omnidirectional sound-deadened room predictions. The new predictions fitted the data overall much better and, in terms of benefit of head rotation, the directional data fitted predictions within 2 dB or so. All benefits of 30-deg head-turns tested were statistically shown to be significant. *t*-tests performed between configuration pairs resulted in *p*-values of 0.03-0.04 for BCIs , $p < 0.01$ for UCIs and $p < 0.001$ for NHs across participants.

CONCLUSIONS

This study demonstrates how a modest and therefore socially acceptable head orientation away from a speaker can provide a significant benefit in understanding speech in noise. The single steady-noise masker situation studied here enables analysis of the fundamental benefit of combining optimum positioning in a room with optimum head orientation, without compromising lip-reading. With the bulk of the noise coming from the side or the rear of the listener, a head orientation of 30 deg is shown to provide a SpIN benefit between 2 and 5 dB for cochlear-implant users without disrupting lip reading. This is a welcome, potentially significant improvement in their challenging speech-in-noise listening situation. Although testing in more reverberant environments and with multiple talker interferers would more realistically mimic a social situation such as a restaurant, this simpler approach demonstrates fundamental benefits. This study also demonstrates how quickly (within minutes after guidance is provided) CI users can learn to reap the benefits a head-orientation strategy can provide. This shows how easily CIs could benefit from simple training. In that, this study has an immediate translational application.

ACKNOWLEDGMENTS

With special thanks to Rosie Lovett and UCL for access to a Chandler House lab.

REFERENCES

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.

Culling, J.F., Jelfs, S., Talbert, A., Grange, J.A, and Backhouse, S.S. (**2012**). “The benefit of bilateral versus unilateral cochlear implantation to speech intelligibility in noise,” Ear Hearing, **33**, 673-682.

Jelfs, S., Culling, J.F., and Lavandier, M. (**2011**). “Revision and validation of a binaural model for speech intelligibility in noise,” Hear. Res., **275**, 96-104.

Lavandier, M., and Culling, J.F. (**2010**). “Prediction of binaural speech intelligibility against noise in rooms,” J. Acoust. Soc. Am., **127**, 387-399.

Litovsky, R.Y., Parkinson, A.J., Arcaroli, J., and Sammeth, C. (**2006**). “Simultaneous bilateral cochlear implantation in adults: A multicenter clinical study,” Ear Hearing, **27**, 714-731.

Loizou, P.C., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., and Roland, P. (**2009**). “Speech recognition by bilateral cochlear implant users in a cocktail-party setting,” J. Acoust. Soc. Am., **125**, 372-383.

Schleich, P., Nopp, P., and D’Haese, P. (**2004**). “Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBI 40/40+ cochlear implant,” Ear Hearing, **25**, 197-204.

Van Hoesel, R., and Tyler, R. (**2003**). “Speech perception, localization, and lateralization with bilateral cochlear implants,” J. Acoust. Soc. Am., **113**, 1617-1630.