Realising the Head-shadow Benefit to Cochlear Implant Users

Jacques André Grange

Thesis submitted to Cardiff University

for the degree of Doctor of Philosophy

February 2015

Declaration

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed .... (candidate) Date...................

Statement 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed .... (candidate) Date...................

Statement 2

This thesis is the result of my own independent work / investigation, except where otherwise stated. Other sources are acknowledged by explicit reference.

Signed .... (candidate) Date...................

Statement 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to other outside organisations.

Signed .... (candidate) Date...................

Statement 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans after expiry of a bar on access previously approved by the Graduate Development Committee.

Signed .... (candidate) Date...................

CONTENTS

List of Figures

List of Tables

List of Symbols

AI	Articulation index	30
AV	Audio-visual	19
BCI	Bilateral CI	76
BE	Better-ear listening	4
BILD	Binaural intensity level difference	9
BMLD	Binaural masking level difference	9
BRIR	Binaural room impulse response	18
BSIM	Binaural speech intelligibility model	57
BTE	Behind-the-ear (CI or HA processor)	70
BU	Binaural unmasking	4
CC	Cross-correlation	47
CF	Characteristic frequency	16
CI	Cochlear implant	2
CIS	Continuous interleaved strategy	15
EBB	Effective binaural bandwidth	61
EC	Equalisation-cancellation	43
EC/SII	the EC/SII model of SRM	57

EE	Excitation-excitation (nerve cell type)	50
EI	Excitation-inhibition (nerve cell type)	51
EM	Energetic masking	2
ERB	Equivalent rectangular bandwidth	57
ESII	Extended SII	39
F0	Fundamental frequency	4
F1	First formant frequency	50
F2	Second formant frequency	50
HI	Hearing impaired	1
HL	Hearing loss	1
HOB	Head-orientation benefit	8
HRIR	Head-related impulse response	71
H_XM_Y	Spatial configuration with head rotated to X°, target at 0° and masker at Y° azimuths	105
HS	Head shadow (effect)	10
IE	Inhibition-excitation (nerve cell type)	51
IEEE	Institute of Electrical and Electronics Engineers (for their "Harvard" sentence corpus)	40
IID	Interaural intensity difference	51
ILD	Interaural level difference	4
IM	Informational masking	2
IPD	Interaural phase difference	9
IR	Impulse response	58
ITD	Interaural time delay	4
jnd	Just-noticeable difference	52
KEMAR	Knowles Electronic Manikin for Acoustic Research	71
LRB	Lip-reading benefit	19
MTF	Modulation transfer function	38

NH	Normally Hearing	1
$N\varphi_nS\varphi_s$	Nomenclature for the binaural presentation of noise and signal with their respective IPDs	43
$N_uS\varphi_s$	- with interaurally uncorrelated noise	43
$N\rho S\varphi_s$	- with a given noise interaural coherence	43
$N\varphi_nS_M$	- with monaural signal presentation	43
OB	Octave band	32
RT60	Reverberation time (to 60 dB attenuation)	35
SE	Standard error (of the means)	92
sEPSM	Speech-based envelope power spectrum model	40
SF-SRT	Speech-facing SRT	59
SII	Speech intelligibility index	35
SNR	Speech-to-noise ratio	2
SNRenv	SNR in the envelope domain	40
SPIN	Speech perception in noise (sentence corpus)	36
SQ	(Binaural) squelch	10
SRM	Spatial release from masking	4
SRT	Speech reception threshold	9
SSEC	Steady-state EC model	62
SSN	Speech-shaped noise (steady state)	10
stBSIM	short-term BSIM	57
STEC	Short-time EC model	62
STI	Speech transmission index	37
STMI	Spectro-temporal modulation index	39
SU	(Binaural) summation	10
TFS	Temporal fine structure	16
TI	Transmission index	38

TIR	Target-to-interferer ratio	66
T_XM_Y	Spatial configuration with target at X° and masker at Y° azimuths	78

ACKNOWLEDGEMENTS

Anne, my beloved sister, thank you for encouraging me to dream away with your old passion for Listening. I too, caught the bug.

Second, but foremost, I would like to warmly thank John Culling for trusting that a materials engineer could convert himself into some sort of hearing specialist! Thank you, John, for your endless support and patience throughout the past three years. I am hugely grateful and indebted to you, as am I to Mathieu and Sam for creating with you such a brilliant model.

Craig, thank you for putting up with my ups and downs over the past few months. I would not have held the course without your support.

Matthieu, I can't imagine ever managing going back to coding without your help. Thank you for your friendship, your moral and technical support.

All the staff at Action on Hearing Loss and AoHL benefactors, thank you so much for funding our research and for your support along the way.

Back at UCL/Chandler's house, Andrew, Rosie, Stuart, Tim, Steve, thank you for your warm welcome and for such a professional attitude.

Val, Hilmar and your teams, your technical support has been essential in Cardiff. Thank you.

Annie, Brice, Marlon, Rob, Barry, sharing that office with you was good, in both fun and studious times. Come see me on the 2nd floor.

Last but not least, Marie, Denis, Liz, Joe, Roy, Brenda, Gill, Mark, Leslie, Pat, Frances, Nora, Fran, Ben, Paul, Sally and Christine, meeting each and every one of you was such a pleasure and gave me the strongest motivation and validation I could ever hope for. Thank you for your generosity, your interest and your ever so valuable feedback.

SUMMARY

Cochlear implant (CI) users struggle to understand speech in noise. They suffer from elevated hearing thresholds and, with practically no binaural unmasking, they rely heavily on better-ear listening and lip reading. Traditional measures of spatial release from masking (SRM) quantify the speech reception threshold (SRT) improvement due to the azimuthal separation of speech and interferers when directly facing the speech source. The Jelfs et al. (2011) model of SRM predicts substantial benefits of orienting the head away from the target speech.

Audio-only and audio-visual (AV) SRTs in normally hearing (NH) listeners and CI users confirmed model predictions of speech-facing SRM and head-orientation benefit (HOB). The lip-reading benefit (LRB) was not disrupted by a modest 30° orientation. When attending to speech with a gradually diminishing speech-to-noise-ratio (SNR), CI users were found to make little spontaneous use of their available HOB. Following a simple instruction to explore their HOB, CI users immediately reached as much as 5 dB lower SNRs. AV speech presentation significantly inhibited head movements (it nearly eradicated CI users' spontaneous head turns), but had a limited impact on the SNRs reached post-instruction, compared to audio-only presentation. NH listeners age-matched to our CI participants made more spontaneous head turns in the free-head experiment but were poorer than CI users at exploiting their HOB post-instruction, despite their exhibiting larger objective HOB. NH listeners' and CI users' LRB measured 3 and 5 dB, respectively.

Our findings both dispel the erroneous beliefs held by CI professionals that facing the speech constitutes an optimal listening strategy (whether for lip-reading or to optimise the use of microphone directionality) and pave the way to obvious translational applications.

1. Introduction

Immanuel Kant's famous quote, "Blindness separates us from things but deafness from people", emphasizes hearing as a prominent social sense among human beings. In most social situations, normally hearing (NH) listeners can follow speech with ease, because their brain separates speech from interfering noise or competing voices. That is, however, not the case for hearing impaired (HI) listeners, even with the help of auditory prostheses. This makes noisy situations very challenging for individuals with hearing loss (HL). When the speaker is visible, the brain's integration of its auditory and visual inputs optimises speech intelligibility in such a way that lip-reading contributes to disentangling the speaker's voice from interferers. Whilst HI listeners seem to get greater benefit from lip-reading than NH listeners (e.g. Grant et al. 1998), lip-reading does not fully compensate for a critical loss of auditory information. Before touching on the impact on speech intelligibility in noise of audio-visual speech integration and sound processing in CIs, one must first introduce the fundamentals of how understanding of speech is affected by competing sound sources.

1.1. Intelligibility of speech in noise

1.1.1. The 'Cocktail Party' problem

In 1953, Colin Cherry was funded by the US air force and naval authorities to investigate a problem air traffic controllers faced at the time, namely that they had the very difficult task of discriminating between a number of pilot voices intermixed over a single loudspeaker. Whilst a lot of fundamental research had been conducted at the time on frequency segregation of pure tones and time separation of clicks, Colin Cherry focussed on speech separation to better understand the recognition of speech in such adverse condition (Cherry 1953). Cherry studied the Cocktail-Party Problem, our ability to focus our auditory attention on a single speaker's voice while filtering out other interfering voices. When asked to separate one of two messages spoken by the same person and presented simultaneously to both ears via headphones, participants reported great difficulty in accomplishing the task, even after several attempts with the same material. Cherry then investigated the impact of dichotically listening to the same two speech streams. Presentation of each stream separately to each ear rendered stream segregation extremely easy, demonstrating that "the process of recognition may apparently be switched to either ear at will". Cherry thought noteworthy to point out that, although our ability to concentrate our attention to one ear and reject what is perceived by the other is not obvious to people who have not experienced hearing tests, “when one tries to follow the conversation of a speaker in a crowded noisy room, the instinctive action is to turn one ear toward him, although this may increase the difference between the ‘messages’ reaching the two ears”.

The purpose of this thesis is to remind ourselves of this instinctive head turn and in doing so, demonstrate how a suitable head orientation strategy can make the difference between a cochlear implant (CI) user being socially isolated or able to engage in and enjoy a conversation when in a noisy social setting such as a restaurant.

1.1.2. The masking of speech

1.1.2.1. Early work

Miller (1947) applied the notion of auditory masking, namely the shift of the threshold of audibility of a sound due to the presence of a masking sound, to speech. He found that the masking of speech is dependent on several characteristics of the masking sound: its energy relative to the speech, its acoustic spectrum and its temporal continuity. Specifically, masking depends primarily on the speech-to-noise ratio (SNR) over the range of frequencies involved in speech, whilst interruptions in the masker decrease its effectiveness. Miller revealed that intensity fluctuations inherent to speech make speech a much less efficient masker than stationary sounds. However when multiple voices were superimposed to generate babble, fluctuations reduced and the babble approximated the masking efficiency of a continuous noise of similar spectral content.

1.1.2.2. Energetic and informational masking

In contrast to Cherry’s work, the masking of speech by noise, as considered early on by Miller, is largely energetic, in that the effectiveness of a noise in masking speech is predictable from the SNR. The effect of energetic masking (EM) on the intelligibility of speech in noise can therefore be modelled with accuracy by considering the energetic ratio between speech and noise in each of the frequency bands relevant to speech. The monaural model of speech intelligibility based on the articulation index (see Section 2.1.1 for details) considers the energetic nature of speech masking by noise. Intensity fluctuations in the noise lead to what was later called *dip-listening*, catching glimpses of the target voice during drops in masking energy. Dip-listening can be modelled as simply due to fluctuations in energetic masking (Rhebergen & Versfeld 2005; Beutelmann et al. 2010).

Informational masking (IM), in contrast, is defined as any masking that is not energetic in nature. It is best illustrated with masking by interfering speech, the linguistic content of which can impair the listener's comprehension of the target speech. This can be considered in terms of the "statistical separation" originally proposed by Cherry (1953). Brungart et al. (2001) concluded that IM was influenced by the similarity/distinctiveness of target and masker voices, after they found that intruding words from the interfering voice were a more frequent source of errors when that voice was the same as the target voice. Evidence that reversed speech could also lead to informational masking (Hawley et al. 2004) suggests that informational masking may not solely occur at a higher linguistic level, but presumably also at a lower phonetic or lexical level, by competing with the target voice in the recruitment of phonetic and lexical processing resources. By extension, any spectro-temporal modulation of a noise that can cause interference with the target recognition at higher linguistic or lower phonetic or lexical levels will lead to informational masking.

1.1.2.3. Predicting masking

This and similar work (French & Steinberg 1947; Fletcher & Galt 1950; Kryter 1962a; Kryter 1962b) led to the formulation and standardisation of the Articulation Index (ANSI 1969), a measure of speech intelligibility primarily used to predict energetic masking in steady-state noise. The AI was later renamed the Speech Intelligibility Index (ANSI 1997). The key assumptions are that the contribution to the SII in a narrow frequency band is, within limits, proportional to the SNR in that band and that the sum contribution of non-overlapping bands can be derived from a weighted sum (see Section 2.1 for a detailed description).

1.1.3. Cues enabling the unmasking of speech in a cocktail party

Hawley et al. (2004) laid out in their introduction the so-far-identified cues that allow NH listeners, in a social setting, to discriminate one speech stream from a number of others and from ambient noise. These cues are described below.

1.1.3.1. Temporal properties of the interfering sound(s)

Continuous noise such as white or coloured noise are effective at masking, because their energetic masking is continuous. Discontinuous noise, modulated noise or interfering voices all enable dip-listening, the listening to portions of the target speech during dips in the masker's energy. Festen & Plomp (1990) explained this as being due to transitory SNR improvements. Top-down processes such as phonemic restoration, a partial or full restoration of masked portions of speech (phonemes) by the brain's statistical analysis of perceived speech, may further contribute to improving intelligibility. Consequently, continuous noise is most effective at masking the target whilst modulated noise, speech-modulated noise, reversed speech and speech alone provide the dip-listening opportunities that increase intelligibility.

1.1.3.2. Fundamental frequencies of target and interfering voice(s)

Interfering voices in a social setting each carry their own fundamental frequency (F0), defined as the lowest frequency of the voice's periodic waveform. Differences between the target's and an interfering voice's F0 enable improved target intelligibility (Brokx & Nooteboom 1982). This occurs through perceptual segregation of signals corresponding with different F0s (Culling & Darwin 1993). The underlying mechanism may be a perceptual cancellation of the interfering voice through inharmonic cancellation (de Cheveigné & McAdams 1995; de Cheveigné 1997; Deroche & Culling 2011; Deroche et al. 2014).

1.1.3.3. Linguistic content of interfering speech and informational masking

Confusion between the linguistic contents of target and interfering speech reduces target intelligibility. This intelligibility reduction is a type of IM since it is not energetic in nature. This effect can be isolated by comparing it to the lesser effect of time-reversed speech that contains all aspects of speech except for intelligible linguistic content.

1.1.3.4. Spatial separation of target and interferer

Making use of the same cues exploited by sound localisation, spatial release from masking (SRM) stems from the reduction in masker effectiveness when it is spatially separated from the target, when compared to the collocated situation. SRM usually involves spatial separation along the azimuthal plane of masker and target sources equidistant from the head. It is defined as the speech reception threshold (SRT) improvement produced by masker separation. It is understood (e.g. Hawley et al. 2004; Culling et al. 2004) as being the result of two separate effects: the head-shadow effect resulting from interaural level differences (ILDs) and the effect of interaural time delays (ITDs). The former contribution to SRM comes from the head obstructing the passage of sound from one side of it to the other. The head-shadow effect is greatest in the high frequency range and reduces the level of a sound source at the ear furthest away from the source, thereby generating interaural level differences (ILDs) between the ears. Its contribution to SRM is made possible by the listener's ability to focus attention on the ear receiving the most favourable SNR, hence it being also referred to as "better ear listening" (BE). The latter, ITD contribution to SRM, also called binaural unmasking

(BU), is due to the longer time taken by a sound wave to reach the ear furthest away from the sound source and the brain's ability to compare the precise timing of the signal's arrival at both ears. BU is most effective in the low to mid frequencies. The trade-off between the two contributions is very similar to that seen in sound localisation. The BE + BU account of SRM is however different to that of auditory scene analysis (Bregman et al. 1990; Bregman 1993) where both ITDs and ILDs contribute to the initial determination of a sound source direction. Edmonds & Culling (2005) manipulated ITDs and ILDs in such a way that one reflected a target azimuth in one hemifield and the other the mirror-image target azimuth in the opposite hemifield. SRM was indistinguishable, whether the two cues coincided or conflicted in terms of target azimuth. SRM was therefore not constrained to the need to focus attention on a particular direction. Thus, despite SRM and localisation making use of the same cues, the authors found evidence against the role of sound localisation in speech intelligibility and showed how ITDs and ILDs make independent contributions to SRM.

Out of the four cues affecting speech intelligibility, we will from now on focus primarily on SRM. The primary interest of this thesis is to demonstrate the basics of how head orientation can be used to optimise SRM for CI users. By making use of a continuous noise spectrally matched to the long-term spectrum of the target speech material, we will restrict our research to SRM alone and not concern ourselves with F0 or dip-listening cues, nor with linguistic content or other forms of informational masking. More complex or ecologically relevant situations can be investigated on the basis of our more fundamental investigation. The following section describes the study of SRM and of its BE and BU contributions over the past 65 years.

1.2. Spatial release from masking

1.2.1. Early studies

Concerned with establishing ways of improving the intelligibility of speech presented to radio operators over earphones in the presence of ambient noise, Licklider (1948) studied the influence of interaural phase relations on the masking of speech by white noise. Taking the monaural intelligibility of speech as a standard for comparison, he found that the advantage of binaural presentation of the speech and the noise depended on the interaural phase relations. Licklider's experiments involved monaural or binaural presentation of speech and white noise over head-phones. When stimuli were presented binaurally, the noise signal was presented in phase, out of phase or incoherently between the ears. Speech was always presented either in phase or out of phase across the ears.

Intelligibility was increased where the interaural relationships of speech and noise differed from one another. The practical application Licklider derived from his results was that radio operator's speech intelligibility in ambient noise could be significantly improved by simply reversing the connections at one of the earphones (creating out-of-phase speech in in-phase noise). A strong correlation between the degree of masking and the extent to which the two sounds were perceived to overlap in phenomenal space suggested that masking and sound localisation involved common neural mechanisms. The idea of a direct relation between masking and localisation however completely broke down in the situation where incoherent noise was employed (and hence noise was perceived as ambient, i.e. not localised), yet masking was reduced. Licklider inferred that central masking (of speech by noise) is not related to the localisation of the speech and the noise in phenomenal space, but to the degree to which the speech and noise share the same interaural relations.

Koenig (1950) was interested in a completely binaural telephone system. He reported some subjective effects of binaural hearing. Two identical microphone-amplifier-earphone lines (forming a binaural system) were used to investigate remote binaural listening, in real time, from one room (the listening room) to an adjacent room (the pick-up room) sound-insulated from the first. In the pick-up room, a dummy head was fitted with the two pick-up microphones instead of ears. Intelligibility of speech in a background of various ecologically relevant noises was compared to a situation with the two receivers fed with a single, common pick-up (a monaural system). Koenig demonstrated that the binaural system was able to "squelch" reverberation and background noises. He proposed that, as the two ears receive signals varying in phase and amplitude for each given sound source, the two signals are combined and compared, either in the nerves or in the brain and that this combination/comparison leads to the squelch effect.

Binaural speech intelligibility with a single interfering noise was most notably investigated as a function of the spatial configuration of the sound sources by Hirsh (1950). In a study entitled "The relation between localization and intelligibility", Hirsh compared monaural and binaural listening and considered the effects of head movements and reverberations. The relationship between interaural phase angle and speech intelligibility demonstrated with speech and noise presented over earphones, suggested that the relative locations of sound sources in free-field would affect intelligibility thresholds. Hirsh measured speech intelligibility thresholds in anechoic and highly reverberant conditions, with speech and thermal noise presented over loudspeakers placed at cardinal positions with respect to the initial head position. SNR measurements at each ear were made with a fixed, dummy head. However, intelligibility thresholds were measured (monaurally and binaurally) with listeners whose *heads were left free to move* in all conditions, thereby confounding the effects of head orientation (unrecorded) and sources separation. Fixed-head thresholds were also acquired, but only in reverberant conditions. Hirsh concluded that, (1) source separation changed the SNR at the ear(s) and hence affected thresholds; (2) head movements appeared responsible for further threshold changes and (3) such further changes are greater in anechoic conditions and with binaural (rather than monaural) listening. In establishing the above, Hirsh had identified the better ear-listening component of SRM and demonstrated that head orientation could lead to further speech intelligibility benefit. Hirsh derived from his findings two factors central to improving the intelligibility of speech in noise and localisation ability of the hearing impaired: first, how important restoring binaural hearing as far as practicable was, second that with a free head, the hearing impaired could reap maximum benefits simply by adopting an optimum head orientation. Hirsh argued that head orientation played a significant role in binaural listening "when the head is allowed to move". An implication identified by Hirsh and directly relevant to this thesis was that for the hearing impaired to "take advantage of these effects, they must have a hearing aid with two separate microphones mounted near the ears and connected each to a separate earphone". Hirsh recommended that bilateral aiding of the hearing impaired was key to improving their speech intelligibility in noise but also that they had to make use of head orientation to reap the full benefit of SRM, thereby clearly setting the scene for this thesis.

Closely following Hirsh's publication, Kock (1950) reported a study on binaural localisation and masking of speech by noise. Kock's study was unique in that he was the first (and last until this thesis) to map out thresholds of speech intelligibility in noise as a function of both (speech and noise) source azimuths *and* head orientation with respect to the speech-facing direction. Kock had set out to test the *cone of confusion* hypothesis illustrated by Figure 1.1.

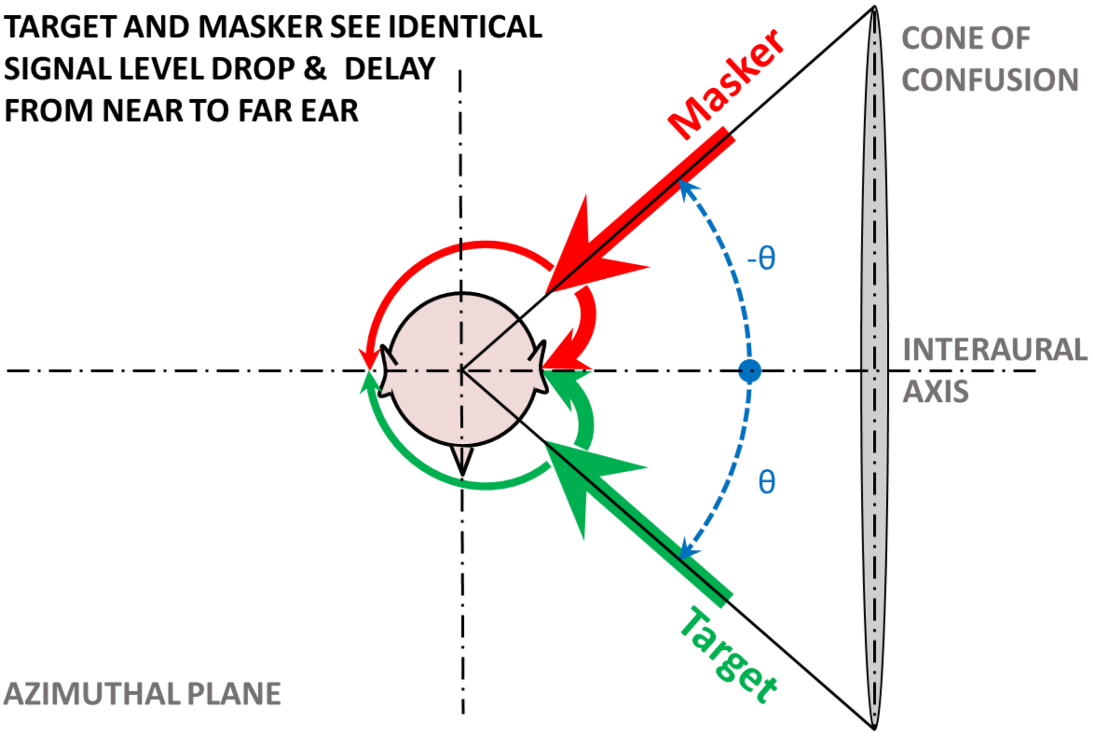

TARGET AND MASKER SEE IDENTICAL SIGNAL LEVEL DROP & DELAY FROM NEAR TO FAR EAR
Masker
CONE OF CONFUSION
-θ
INTERAURAL AXIS
θ
Target
AZIMUTHAL PLANE

Figure 1.1: The cone of confusion intersects the azimuthal plane along two azimuths. If the speech comes from θº away from the interaural axis, azimuthally separated noise masks the speech best when it comes from -θº. Both then share the same ITD and ILD

When one approximates the head and ears to a sphere with two diametrically opposite sound-pressure pick-up points on its surface, sounds roughly share the same ITDs when coming from any direction along the surface of a cone, the apex of which is at the centre of the head and the axis of which is the line between the two ears. A direct consequence was that Kock found speech intelligibility to be worst when speech and noise came from locations on the same cone of confusion, whether speech and noise were collocated or not. A fixed-level standard sentence repeated in a loop was presented at 0º and masked by a thermal noise presented at 90, 180 or 270º, the level of which was adjustable by the listeners. At head orientations between 0 and 360º and every 45º, the listeners adjusted the noise level to reach a set threshold in perceptibility of the speech. Just as predicted, the *cone of confusion* hypothesis was accepted when in all configurations, speech discrimination failed when the head orientation was such that speech and noise directions lay on the same cone of confusion. In between minima and at optimum head orientations, SRM peaked 12-15 dB above the lowest minimum in each configuration. The potential for HOB was maximum with speech and noise at 0º and 180º respectively. We invite the reader to ponder on Kock’s findings since we will later report SRM measurements in all three spatial configurations chosen in his study. Kock inferred from his findings that the brain preserves and evaluates time delays to achieve both directional localisation and discrimination of speech against reverberation and background noise. This was the original idea that formed a basis for the Equalisation-Cancellation theory and model of binaural unmasking discussed in Section 2.2.2.

1.2.2. Towards a refined understanding of SRM

1.2.2.1. Definitions and acronyms

A few definitions are required at this point, to assist the reader.

- *Detection threshold*: the signal level at which the listener, under a given set of conditions, and maintaining a fixed criterion, correctly reports whether the signal is present a given percentage of the time. The commonly used 2-up 1-down adaptive threshold rule gives the 70.7% threshold.

- *Speech-to-noise ratio (SNR)*: The ratio between speech and noise levels (in dB). Unless specified otherwise it refers to the SNR at source.

- *Speech reception threshold (SRT_P)*: the level (when measured in quiet) or SNR (when measured in noise) at which the listener correctly identifies a given percentage of the words correctly. Unless stated otherwise, SRT refers to 50% intelligibility.

- *Binaural masking level difference (BMLD)*: the difference in detection thresholds (in dB) between a reference diotic or monotic condition and any other binaural condition.

- *Binaural intelligibility level difference (BILD)*: the difference in SRT between a reference diotic or monotic condition and any other binaural condition. Unless stated otherwise, 50% intelligibility is implied.

- *Interaural level difference (ILD)*: the difference in signal level between the ears.

- *Interaural time delay (ITD)*: the difference in signal arrival time between the ears.

- *Interaural phase difference (IPD)*: the difference in signal phase between the ears.

- *Better-ear listening (BE)*: the ability of a listener to attend to the ear receiving the better SNR within each of the frequency bands analysed by the cochlea. Hence BE only occurs when ILDs are present.

- *Binaural unmasking (BU)*: the ability of the brain to achieve improved noise rejection via processing of ITDs.

- *Spatial release from masking (SRM)*: the improvement in SRT (in dB) from a collocated-in-front to a separated speech-and-masker situation, all other parameters remaining the same. Note SRM can go negative in monaural conditions when the masker and/or the target are respectively moved to the same and/or opposite side of the median plane as the operating ear.

- *Summation (SU)*: The benefit (in dB) of adding the second or poorer ear to the intelligibility of speech when the noise is collocated with the speech, in other words when there are no ITD or ILD differences between speech and noise.

- *Squelch (SQ)*: The benefit (in dB) of adding the second or poorer ear to the intelligibility of speech when the noise is separated from the speech, in other words when ITDs and ILDs differ between speech and noise.

- *Head-shadow effect (HS)*: Typically defined as the monaural SRT improvement (in dB) when the noise is moved from an azimuth in the hemifield on the side of the operating ear to the mirror-image azimuth in the opposite hemifield. HS is sometimes defined as the benefit of adding the acoustically favoured ear. Note that the two definitions are equivalent as long as BU and SQ do *not* occur.

- *Speech-shaped noise (SSN)*: continuous noise made to share the same long-term spectrum as the speech material (the target material unless stated otherwise).

1.2.2.2. SRM for NH listeners

A number of early studies (Kock 1950; Schubert 1956; Levitt & Rabiner 1967a; Carhart et al. 1967) focussed on IPDs and ITDs alone. As reported by Levitt & Rabiner (1967a), binaural release from masking using IPDs over headphones (detection of out-of-phase tones in in-phase noise) led to BMLDs of up to 13 dB, but BILDs only up to 6 dB. When manipulating the ITDs instead of IPDs, large ITDs (0.5 to 10 ms) led to BMLDs of up to 12 dB, but BILDs plateauing at only 3 dB.

Later studies considered SRM and therefore included the head-shadow effect. Dirks and Wilson (1969) studied the effect of the spatial separation of sound sources on the intelligibility of spondaic or PB words in noise and synthetic sentences masked by noise or competing speech. Artificial-head recordings were presented over headphones or listeners were placed in a sound field. Consistent with previous studies focussing on ITD alone, speaker locations giving rise to different ITDs for speech and noise led to the higher (binaural) intelligibility. In monaural conditions, loudspeaker positions that led to a higher SNR (than that for a position ipsilateral to the operating ear) gave rise to higher intelligibility. The results also showed that intelligibility was superior in binaural compared to (better ear) monaural presentation only when ITDs were present. The monaural outcomes were early evidence of the BE contribution to SRM, whilst the superiority of binaural outcomes over near-ear monaural outcomes only when ITDs were present constituted early evidence of the BU contribution to SRM. However, the roles of ITDs and ILDs were not clearly disentangled.

In order to gather an insight into everyday listening situations, Plomp (1976) set out to measure SRM with connected discourse masked by SSN or other connected discourse presented over loudspeakers in an anechoic room or a reverberant room. The reverberation time of the room was adjusted with sound-absorbent panels. The target speech was always in front of the listener (as Plomp assumed this to be the most natural or relevant situation) whilst the masker was presented collocated with the target or separated from it by up to 180º (in 45º steps). SRTs were measured binaurally or monaurally, following the Békésy up/down tracking technique (LeZak et al. 1964), with the target speech level adjusted by the listener to make the speech 'just intelligible' in a fixed masker level. SRM increased with increased masker separation up to a maximum of 6 dB in anechoic conditions (at ±135º of separation), thereafter reducing to 3 dB (at 180º). SRTs were typically 3 dB lower with competing speech (due to dip-listening). Reverberation had a two-fold detrimental effect on intelligibility by both worsening SRTs and reducing SRM (maximum SRM gradually dropping from 6 dB to 1-1.5 dB with reverberation time, RT_{60}, increasing up to 2.3 s). Binaural SRTs were 2.5 dB higher than monaural SRTs irrespective of reverberation or source separation. Plomp concluded that in everyday situations NH listeners easily separate the target speech from competing speech when the reverberation time is below 0.5 s. Results however suggested that unilateral deafness would significantly reduce intelligibility when listeners do not benefit from the head shadow effect (i.e. the interferer is on the same side as the normally-hearing ear). Since patients with sensorineural hearing loss typically required 5 to 15 dB larger SNR than NH listeners, his results led Plomp to recommend binaural hearing aids with high quality directional microphones in a low reverberation setting. The Békésy up/down tracking technique was later identified as suffering from the listener's inability to completely ignore differences in the loudness of speech. Following the validation of an unbiased and more accurate adaptive SRT measurement method (Plomp & Mimpen 1979) involving sets of sentences, Plomp and Mimpen (1981) demonstrated that SRM in anechoic conditions could be as high as 10 dB (with the masker at 112.5º). This was in good agreement with Platte & Vom Hövel's (1980) findings (making use of numbers as speech material).

Bronkhorst and Plomp (1988) pointed out that difficulties arise when relating BILDs observed with ITDs to those obtained in free-field conditions. ITD effects in the free-field might be influenced by ILD effects. Moreover, although the dependence of

ITDs on frequency, caused by the diffraction of sound pressure waves by a human head, had been precisely modelled in the azimuthal plane by Kuhn (1977), it was unknown to what extent BILDs were affected by that phenomenon. The authors therefore embarked on a ground-breaking study aimed at disentangling the contributions of ITDs and ILDs to SRM, thereby establishing a much improved understanding of SRM and paving the way to the refinement of models of SRM. In the main study, the authors simulated a free-field situation by presenting KEMAR-manikin recordings of meaningful sentences over headphones and measuring 50% intelligibility SRTs with the adaptive method of Plomp and Mimpen (1979). Speech was always presented in front of the listener (0º azimuth) and hence did not lead to ITDs or ILDs. SSN was simulated as presented from 0 to 180º azimuth every 30º, either as original, free-field recordings (FF noise) or FFT-processed so as to contain only ITDs (dT noise) or only ILDs (dL noise). SRM was computed for each of the FF, dT and dL noise cases as the spatially-separated-condition SRT subtracted from the collocated-condition SRT. The resulting $\mathrm{SRM}_{\mathrm{FF}}$, $\mathrm{SRM}_{\mathrm{dT}}$ and $\mathrm{SRM}_{\mathrm{dL}}$ were then compared as a function of the noise azimuth. Whilst $\mathrm{SRM}_{\mathrm{dT}}$ was found to be around 5 dB and independent of noise azimuth between 30º and 150º, $\mathrm{SRM}_{\mathrm{FF}}$ and $\mathrm{SRM}_{\mathrm{dL}}$ shared the same trend that resembled an inverted parabola peaking at 90º noise azimuth (at 10 and 8 dB respectively). Interestingly, the sum of $\mathrm{SRM}_{\mathrm{dT}}$ and $\mathrm{SRM}_{\mathrm{dL}}$ exceeded $\mathrm{SRM}_{\mathrm{FF}}$, indicating ITD and ILD contributions to SRM were not completely independent.

Concerning the advantage of binaural over monaural hearing, many initially thought that ITDs were the primary cue for the binaural system. The results of Carhart et al. (1967) showed that ITDs yield only a moderate contribution to SRM. Greater BU is accomplished when only detection rather than intelligibility (of speech in noise) is measured. This was demonstrated by Levitt and Rabiner (1967a), who found BMLDs up to 10 dB larger than BILDs. In addition, it was suggested by Carhart et al. that ILDs have a degrading effect on BU. This is supported by the results of Bronkhorst & Plomp presented in the previous paragraph. There, the contribution to SRM by ITDs, $\mathrm{SRM}_{\mathrm{dT}}$ is consistently larger than the difference between $\mathrm{SRM}_{\mathrm{FF}}$ and $\mathrm{SRM}_{\mathrm{dL}}$, suggesting that their role is smaller when the two cues are combined. Two key findings by Bronkhorst & Plomp were therefore that, (1) ILDs, were the primary cue for the binaural system, not ITDs and (2) ILDs decreased the effect of ITDs (or BU). The latter finding can be explained by the fact that in a sound field, when ILDs are large, the speech and noise signal levels are so different between the ears that the ITDs are less accurately compensated for by the auditory brain and contribute less to SRM. Artificially removing ILDs therefore has for an effect to increase the effect of ITDs.

Bronkhorst and Plomp (1992) later investigated in a free-field simulation the effect on SRM of distributing multiple maskers azimuthally with speech in front and up to 6 speech-modulated speech-shaped noises. This aimed at mimicking a cocktail party but excluded informational masking, the effects of F0 differences or reverberation. As SRTs were measured both binaurally and monaurally for symmetrical and asymmetrical spatial distributions of maskers, the monaural contribution to SRM could be separated from the binaural advantage. Constant BU was found of approximately 3 dB, whilst results with a single masker at 90º showed a considerably larger monaural BE contribution of 8 dB. The BU was half of that found with ITDs alone, confirming that with large ILDs, the ITD contribution is reduced. The authors noted that at equal 'voice' loudness, speech was intelligible for NH listeners with up to 6 spatially distributed interferers whilst hearing-impaired listeners (sensorineural hearing loss) struggled to cope from 4 interferers upward as they seemed unable to exploit dip-listening.

In another free-field simulation, Peissig and Kollmeier (1997) further investigated the 'effective' directionality of binaural unmasking with up to three interferers, either SSN (created by superimposing many voice samples) or individual voices. With speech in front, up to two interferers were fixed at the azimuths that led to maximum SRM for a single masker (105 and 255º). SRT was measured as a function of the azimuth of an additional interferer. The authors interpreted their data as evidence that NH listeners suppress interferers from one azimuth at a time. Differences in the pattern of data for SSN and speech interferers led them to conclude that listeners could utilize dips in one interferer to suppress another, spatially separated interferer. They showed that SRM decreased rapidly as the number of interferers increased and as their azimuthal separation increased. What the authors did not highlight although it was probably most visible in their data set (as well as in Müller 1992) was that with a single SSN, a dip in SRM was quite visible at 90 or 270º of separation.

In proposing clinical tests of binaural hearing, both Bronkhorst and Plomp (1990) and Koehnke & Besing (1996) favoured testing with reverberation (typically 1s reverberation time) with the listener facing the speech and noise collocated or separated from the speech by 90 or 270º. The tests were proposed in real and virtual environments respectively. In both studies SRM was found to be only 4 dB, as opposed to up to 10 dB in anechoic conditions. The spatial configuration chosen here had become a standard for most studies and clinical tests of SRM. Yet, such spatial configurations present two issues: firstly, as was highlighted much earlier by both Hirsh (1950) and Kock (1950) but never re-visited since, head orientation away from the speaker could lead to much improved SRM; secondly, both Müller (1992) and Peissig and Kollmeier (1997) found that SRM was reduced at a 90º noise separation. This 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 & Martens 1998). The first issue highlights that the standard configuration with the speech in front does not maximise the SRM potentially available, thereby reducing the dynamic range of the test, the second that SRTs measured with the masker at 90º are also suboptimal due to the bright spot. Moreover, because the bright spot is very small the measurement may be subject to a high variability due to its dependence on the exact head orientation. Such variability in SRM outcomes is obvious in the review paper by Bronkhorst (2000), where anechoic SRM from various studies and in the same spatial configuration ranged from 6 dB to 10 dB, most of the discrepancy between studies being plausibly attributable to inaccuracies in the exact head orientation adopted by listeners or, for virtual presentations, in the positioning of the artificial head during material or impulse response recording.

At this point, we invite the reader to note that, with the exception of Hirsh (1950) and Kock (1950), all studies reviewed above placed the target speech in front of the listener. This was originally motivated by facing the speech being considered a more natural listening attitude (Plomp 1986). However, in adopting such a ‘standard’, most studies do not demonstrate the maximum obtainable SRM resulting from a combination of masker separation and optimal head orientation.

1.2.2.3. Effect of sensorineural hearing loss on SRM

In studying SRM, apart from the original motivation driven by improvements in telecommunications and researchers’ endeavour to reach a more fundamental understanding of normal hearing, an overwhelming drive in most of the studies reviewed in the previous section stemmed from HI people experiencing great difficulties with spatial hearing in noisy environments, difficulties that could not be simply accounted for by their audiograms. The shift in their audibility thresholds alone was indeed not enough to explain their reduced speech-in-noise intelligibility (see the articulation index model in Chapter 2). Plomp (1976) highlighted that unilateral deafness can severely impede speech intelligibility in noise when the noise is in the same hemifield as the normally operating ear, and particularly so in anechoic conditions. Plomp further showed that a larger handicap was found with sensorineural HL, where listeners could exhibit SRTs in excess of 5 dB worse than NH listeners. Bronkhorst and Plomp (1992) repeated this finding by showing that with moderate HL, SRTs were 4 to 10 dB higher than for NH listeners and that only about a third of the SRT variance with HL could be accounted for by pure-tone audiograms. They further showed that, whilst NH listeners could benefit from dip-listening, HI listeners benefitted little, which led to them being able to cope with far fewer spatially distributed voice-like interferers than NH listeners. The authors attributed the reduction in the HI dip-listening benefit to a combination of effects: (1) the advantage during dips in an interferer is reduced by threshold elevation, which results in a narrowing of their available dynamic range, limiting the information provided by quieter portions of the speech audible by NH listeners; (2) the reduced temporal resolution usually associated with hearing impairment (Elliott 1975) directly reduces HI listeners' dip-listening benefit; and (3) the comodulation masking release (Hall & Grose 1988) is reduced in HI listeners, which will also directly reduce the dip-listening benefit. These findings were echoed by Peissig & Kollmeier (1997) who reported that sensorineural HL led to increased SRTs and reduced SRM, possibly resulting from a reduced ability to exploit dip-listening. Listeners with sensorineural HL exhibited a reduction in the BE contribution to their SRM. In a simulation of hearing impairments, Bronkhorst & Plomp (1988) had previously extended their main study (discussed in the previous section) to explore ITD and ILD contributions to SRM with attenuation of 20 dB in the right or left ear. Their results mirrored Peissig & Kollmeier's (1997) in producing a BE reduction of up to 3.5 dB. In addition, they found that the BU contribution to SRM was also significantly reduced (by up to 2 dB) as ITDs could not be exploited as effectively. This demonstrated how HI listeners could see their BU benefit almost halved and their BE benefit reduced by a third (in dB terms) in the spatial arrangements least favourable to each SRM contribution.

1.2.2.4. SRM for CI users

Cochlear implants transduce sound over an array of 12 to 22 electrodes, each fed with an independently processed, band-filtered signal. Once the implant is inserted in the cochlea, sound information is electrically coupled from the electrode array to the CI user's spiral ganglia. The electric field generated by electrical stimulation spreads along the spiral ganglia and current can flow between electrodes. As a result, early multiple-electrode CI trials attempting to excite neurons in an analogue manner were not successful as the speech information carried by each electrode (or spectral band) could leak to other electrodes, causing interference between bands and rendering speech unintelligible. The most successful countermeasure was to stagger in time the excitation of different electrodes along the cochlea by making use of staggered pulses that reproduce the envelope of the signal within each frequency band. This continuous interleaved strategy (CIS), widely used today, helps restore speech intelligibility to a level approaching that of NH listeners at a normal speech level and without noise. This is the case provided that (1) patients have sufficient nerve survival along their spiral ganglia, (2) coupling of the pulses into these nerves is efficient and (3) neural pathways to the brain have not suffered too much degeneration over the period of deafness preceding implantation.

All being well, CI users should have their high frequency hearing sufficiently restored that they do not suffer from the loss of BE contribution to SRM that listeners with sensorineural HL typically experience. In that sense, unilateral CI users are expected to benefit from the head-shadow effect, provided that the spatial distribution of sounds favours their implanted side. Unilateral users (with little to no hearing in the other ear) cannot by definition benefit from BU. For bilateral CI users, because the temporal fine structure (TFS) of sounds is lost by most current sound processing strategies, ITDs are not accurately transmitted to the brain and as a result little or no BU can occur. Furthermore, each point along the basilar membrane is tuned to a given characteristic frequency (CF). Even if ITDs could be exploited at low frequency, it is very unlikely that the electrode arrays would be inserted to exactly the same depth in the two ears, so there would be a misalignment across ears. The neurons excited maximally by the nth electrode in one ear would have different CFs from the neurons maximally excited by the nth electrode in the other ear. This means that the normal process for measuring interaural time delays may be compromised. Researchers compensate for insertion depth discrepancies or differing left/right CI electrode arrays by pitch-matching electrode pairs between implants (e.g. Van Hoesel & Tyler 2003; Dorman et al. 2007; van Hoesel et al. 2009; Litovsky et al. 2010). Signs of pitch perception adaptation have been found (e.g. Reiss et al. 2011), but only at very low frequencies which are nearly irrelevant to speech perception has BU been observed in CI users, and only with narrow-band maskers (Long et al. 2003; Van Hoesel 2004). Even combining pitch matching and advanced coding strategies aimed at restoring faithful ITDs at low frequency, Van Hoesel et al. (2008) found no binaural advantage. Overall, very little evidence has been found to date of ITDs helping CI users' spatial hearing (including localisation, see Aronoff et al. 2010). Although lower pulse rates have been shown to facilitate ITD discrimination, higher pulse rates provide better speech recognition. The trade-off between better speech recognition at high pulse rate and better pulse timing sensitivity (or faithful ITD delivery) at low pulse rate was recognised as a conundrum by Churchill et al. (2014). The assumption that negligible BU occurs in bilateral CI users is therefore reasonable and will be made hereafter (see Chapters 2 and 5).

1.2.2.5. ‘Traditional’ measures of SRM in CI users: speech-facing SRM

A search of the literature revealed eighteen studies involving CI users where SRM and/or binaural advantage were considered in various spatial configurations. These studies compared bilateral to unilateral (with one CI disabled) outcomes. All but three (Laszig et al. 2004; Laske et al. 2009; Culling et al. 2012) considered only the traditional, speech-facing situations. The benefit of SRM was always shown by contrasting the collocated-masker situation against separated situations with the masker at 90 or 270º, thereby leading to suboptimal SRM and including the bright spot issue discussed above. Of these 15 studies, 8 (Müller et al. 2002; Tyler et al. 2002; Peters et al. 2007; Tyler et al. 2007; Buss et al. 2008; Litovsky et al. 2009; Eapen et al. 2009) reported their results as percent correct. Because we are mostly interested in SRM in terms of improvement in SRT and percent-correct results cannot be converted to SRTs without making assumptions about the underlying psychometric function, we will focus on the remaining 7 studies (Van Hoesel & Tyler 2003; Schleich et al. 2004; Litovsky et al. 2006; Loizou et al. 2009; Van Deun et al. 2010; Lovett et al. 2010; Murphy et al. 2011). These studies reported a mix of SRM and measures of head-shadow effect (the definition of which varies slightly across publications) with measures of binaural summation and/or squelch. The HS measure was typically defined as the monaural SRT improvement when the noise was moved from ipsilateral to contralateral to the enabled CI and was sometimes defined as the benefit of adding the acoustically favoured ear (which matches the first definition if no binaural unmasking exists, see e.g. Litovsky et al. 2006). SRM ranged from 3.5 to 5 dB and HS from 4.5 to 7 dB. SU and SQ were reported to be below 1 dB (and rarely significant) and between 0 and 2 dB (with varied significance) respectively. With the exception of Culling et al. (2012), none of the studies provided a measure of the level of reverberation in the test room, often reporting the use of sound-deadened audiology suites or loosely referring to an anechoic room. As we will see later, reverberation can have a significant impact on SRM and should therefore be more tightly controlled and reported.

1.2.2.6. Evidence of the additional head-orientation benefit to CI users

Three studies (Laszig et al. 2004; Laske et al. 2009; Culling et al. 2012) investigated SRM and/or HS in non-speech-facing configurations, with target and masker separated symmetrically about the 0º azimuth by ±45 or ±60º. In this case, HS is measured as unilateral SRT improvements from the unfavourable to the favourable situation (mirror-imaged about the median plane). As highlighted in Culling et al. (2012) HS and SRM were predicted by the Jelfs et al. (2011) model of SRM (see Chapter 2) to be maximum around ±60º of symmetrical separation. Culling et al. measured HS for unilateral CI users at 18 dB for ±60º of separation. This matched model predictions, but was substantially more than the 10-11 dB separately reported by Laszig et al. and Laske et al. for ±45º, where the model predicted 16 dB. After eliminating the possibility of speech material or (limited) directionality of CI microphones being responsible for such a discrepancy, Culling et al. highlighted that their study was unique in that they had acquired impulse responses in the test room and derived model predictions from them. Thus, the variability in reverberation between test rooms was plausibly the best explanation for the discrepancy observed.

Culling et al. were the first since Kock (1950) and Hirsh (1950) to highlight, that restricting SRM measurements to speech-facing or symmetrical separation situations would not reveal the full SRM potential. Making the assumption that bilateral CI users' implants are equally effective for reception of speech in noise, the large HS measured with unilateral CI users enabled Culling et al. to argue that the benefit of bilateral over unilateral implantation was much larger than had been previously reported. Knowing that postoperative speech-receptive capacities of CIs vary widely, the authors assessed the impact of unequally effective CIs by using SRTs from the 6 bilateral CI users tested by Loizou et al. (2009). These showed interaural inequalities statistically comparable to another 34 subjects tested by Litovsky et al. (2006). However, when the model was adjusted to reflect individuals' interaural inequalities, the correlation between model predictions and the Loizou et al. (2009) data was reduced slightly from 0.97 to 0.96. This supported the assumption of equally effective CIs as being reasonable.

1.2.2.7. Expected effect of reverberation on SRM

In the event of diffuse noise (typically due to high levels of reverberation) or noises coming from multiple azimuths on both sides of the median plane, both unilateral and bilateral CI users will experience great difficulty in following speech. This was shown by Loizou et al. (2009) in a speech-facing situation with maskers distributed at -30, 60 and 90º. In that configuration, not only did thresholds increase typically by 6 dB, SRM was also reduced to 2 dB. Misurelli & Litovsky (2012) also measured SRM below 1 dB when two maskers were placed at +90 and -90º.

The question remains as to the impact reverberation has on SRM measured with a single masker. Most studies report single-masker SRM measurements being acquired in (supposedly) anechoic or sound-deadened rooms such as audiology suites. Culling et al. (2012) made use of an acoustic ray-tracing model to generate binaural room impulse responses (BRIRs). These were in turn fed into the model of SRM to predict the effects of different rooms and different spatial configurations. SRM (in dB) was predicted to increase nearly linearly with the absorption coefficient of a virtual room.

Reverberation time is very widely used as a measure of reverberation (RT_{60} being the time it takes for the reverberated sound energy to fall 60 dB below the direct sound's energy). By feeding into the model BRIRs acquired in a variety of rooms with widely ranging reverberation (for a source 1.5 m away from the head), Culling et al. demonstrated that no relationship could be found between RT_{60} and SRM predictions. Instead, the authors proposed the direct-to-reverberant energy ratio of the noise as being an adequate predictor of SRM. For the rooms acoustically measured, SRM (in dB) was predicted to drop when compared to a strictly anechoic environment by 10 % for typical sound-deadened test rooms, by 40 to 50% in a typical living environment (such as a living room) and by up to 80% in confined, highly reverberant environments such as a stairwell. For a single interfering source, highly reverberant environments of large dimensions such as a typical cafeteria were predicted to lead to only 20% drop in SRM. However, large social settings tend to be filled with many interfering talkers that bring up the overall noise level and significantly reduce SRM. It is also worth noting that reverberation would be much more disruptive for a distant noise source, compared to the setup described above for BRIR measurements. These considerations highlight the importance to CI users of the management of reverberation in social settings.

1.3. Visual cues to speech intelligibility & audio-visual integration in speech perception

In determining the potential head-orientation benefit to CI users when listening to speech in noise, one must take into account the impact of the speaker being visible to the listener. Indeed, lip-reading plays an important role in a HI listener's recognition of speech. In order to better appreciate the measures of lip-reading benefit (LRB) reported in Chapters 4 and 5, the following sections provide a context by reviewing the current scientific understanding of audio-visual (AV) speech perception as well as the reported measurements of LRB.

1.3.1. Lip-reading in NH listeners

Since the 1950's, the benefits to NH listeners of AV over audio-only presentation of speech have been observed in terms of speech detection (e.g. Repp et al. 1992; Grant & Seitz 2000; Bernstein et al. 2004) and/or recognition (e.g. Sumby & Pollack 1954;

Erber 1969; MacLeod & Summerfield 1987; Summerfield 1987; Middelweerd & Plomp 1987; Macleod & Summerfield 1990) in quiet or in noise.

1.3.1.1. Quantifying the benefit of lip-reading

Making use of a percent-correct identification measure for spondaic words, Sumby & Pollack (1954) demonstrated that the visual contribution to the intelligibility of audio-visual speech was greater at low SNR. They measured LRBs ranging 5-22 dB depending on the size of the response set. These findings were echoed by Erber (1969) who found LRBs of 5-10 dB. Thus, early studies found that the poorer the audio-alone performance, the greater the LRB. However, studies using percent-correct measures failed to find a link between LRB and individual differences in lip-reading ability. This was explained by Macleod & Summerfield (1987) as due to ceiling effects. The authors showed that the use of an adaptive measure of SRT removed the ceiling effect. They carefully selected and ranked BKB sentences (Bench et al. 1979) according their lip-reading difficulty. Macleod & Summerfield reported a positive correlation between visual-alone performance and LRB (ranging 6-15 dB) and between sentence lip-reading ease and LRB (ranging 3-21 dB). Middelweerd & Plomp (1987) measured LRB for the intelligibility of sentences (Plomp & Mimpen 1979a) in SSN at 4.6 dB for young adults and 4 dB for elderly listeners. Macleod & Summerfield (1990) later refined their measure of SRT by modifying the Plomp & Mimpen (1979) adaptive method for audio-only and AV SRT measurements in noise. Making use of new sentence lists with balanced lip-reading difficulty, the authors measured on average a 6.4 dB LRB.

1.3.1.2. The importance of AV synchrony

McGrath and Summerfield (1985) studied the effect of intermodal time relations on AV speech recognition. Their aim was to determine what was the minimum asynchrony that disrupts the AV intelligibility of speech, a measure important in the design of sound processing algorithms for hearing aids and cochlear implants, since any sound processing aimed at enhancing the acoustical speech signal delays it. When replacing the audio signal with pulse trains that conveyed the fundamental frequency (F0) of the talker’s voice, soundtrack delays up to 80 ms had little effect on NH listener group mean performance. However, a sub-group of superior lip-readers showed gradually improving performance as the delay was reduced from 80 to 0 ms. A second experiment showed that NH listeners, whether good or poor lip-readers, possess insufficient sensitivity to AV onset asynchrony for using natural intermodal onset timing cues (at a minimum 30 ms) in phonemic identification. The combined results of both experiments led the authors to conclude that acoustic signal delays up to 40 ms should not materially affect AV speech recognition. This is also important to our Chapters 4 and 5 audio vs. AV experimental design, as any on-the-fly manipulation of the soundtrack must maintain sufficient AV synchrony to not compromise AV integration.

1.3.1.3. What underlies the benefit of lip-reading

The review of lip-reading and audio-visual speech perception by Summerfield (1992) presents an overview of the understanding of lip-reading and AV speech reception. Lip-reading is defined as the perception of speech by purely visual observation of the talker's articulatory gestures. The lips, teeth and tongue are the visible parts of the vocal tract. The internal shape of the mouth and lips acts as a filter to the uniform acoustic spectrum generated by the vibrating vocal folds in the larynx. This can be considered as a filter function characterised by the patterns of peaks and troughs imposed on a uniform spectrum. The resulting changes in resonant frequencies convey, amongst other phonetic aspects of speech, the identity of vowels and the place or articulation of consonants. Consonants are produced via rapid articulatory gestures and are conveyed by correspondingly rapid changes in resonance. Vowels, the result of slower, longer movements are more static acoustically. The fine spectral detail that reflects the place of articulation of consonants in the mid to high frequencies is more severely affected by noise (Miller & Nicely 1955), reverberation or distortions such as those resulting from sensorineural HL (Walden et al. 1975). Prosody and phonetic features, such as nasality and voicing of consonants (gross time-amplitude patterns at low frequencies) or periodicity conveying the intonation contour of connected speech, are more robust in adverse listening conditions. The velum and larynx that produce such robust features are not visible, while the lips, teeth and tongue responsible for the least robust features are. It is therefore easy to see how lip-reading can complement audition in adverse conditions.

The question remains as to how the brain integrates converging, bimodal information. When considering the degree to which acoustic speech signal and lip movements share spatial and temporal properties, Summerfield (1987) proposed two possible contributions of visual cues to AV speech understanding in noise: (1) segmental (e.g. consonants, vowels) and supra-segmental (e.g. intonation, stress, rhythmic pattering) information that is *redundant* with acoustic cues; (2) segmental and supra-segmental information that is *complementary* to acoustic cues when acoustic cues are masked or distorted. Voice pitch cues are known to provide important segmental and supra-segmental information usually invisible and therefore complementary to the visual input (Rosen et al. 1981; Grant 1987). Changes in the area of the lip opening correlate with the speech signal envelope and provide a reduction in onset uncertainty of syllables and words. An extension of Summerfield's work reported a correlation between lip kinematics and the second and third formant frequencies (Grant & Seitz 2000), which is coherent with lip-readers' ability to extract primarily place of articulation information.

The audio and visual inputs are bound to the same articulatory event. This co-modulation across modalities suggests both inputs should be processed together. A useful way of analysing AV integration is to present conflicting auditory and visual stimuli and see which one dominates or if the AV percept lies somewhere between the auditory and AV stimuli. McGurk & MacDonald (1976) reported how, when repeated utterances of the syllable [ba] were dubbed on to lip movements for [ga], normal adults reported hearing [da]. This is an instance of the 'McGurk effect', the bimodal fusion of conflicting speech stimuli. At least three facts emerge from these findings: (1) integration is not simple averaging, (2) the phonetic changes due to conflicting visual cues are not large in acoustic terms, and (3) incompatibility between auditory and visual inputs is generally not detected. Summerfield & MacGrath (1984) showed that perceptual shifts due to conflicting visual cues also occurred with vowels, with the AV incompatibility much more often detected and with vowel identification biased toward the visual input. Observers behaved as if they computed a continuous estimate of the vocal tract filter function from the evidence of both modalities. Summerfield & McGrath (1984) inferred from this (1) that AV integration has to occur before speech sounds are phonetically classified, (2) that the common metric of integration is an auditory representation of the filter function (in that observers report that they 'hear' the impact of AV integration), and (3) that visual evidence of the filter function is obtained by a process of computation, e.g. a heuristic strategy akin to a look-up table that takes account of the redundancy between lip-shape and tongue position.

The visual cues alert the listener to the temporal and possibly spatial locations of the speech stimulus in such a way that congruent information may direct the auditory attention. Brain imaging techniques have revealed how the auditory cortex is influenced by lip-reading (Sams et al. 1991; Calvert et al. 1997), but also how the nervous system is sensitive to converging input from multiple senses (at least the superior colliculus, see e.g. Meredith et al. 1987; Wallace et al. 1993). Most models of speech perception AV integration however assume independent auditory and visual information sources. The more recent model by Massaro (1998), the fuzzy logical model of perception (FLMP), includes a third, independent cross-modal source of information integrated in a multiplicative manner. FLMP is essentially based on the multiplication of probability of auditory and visual pattern recognition and matching of the two identified patterns. The ways Summerfield and Massaro apprehend AV integration of speech are orthogonal to each other, Massaro's exemplifying the 'generalist' or 'engineer' angle and Summefield the 'specialist' or 'researcher' approach. The two authors conflict in that Massaro's assumption that AV integration is just another instance of general cognitive information integration (here at a phoneme level) is rejected by Summerfield's demonstration that speech integration has to occur before phonetic classification of sounds. They both, however, have something to contribute.

1.3.1.4. Lip-reading benefit in complex listening situations

A study by Helfer & Freyman (2005) investigated the role of lip-reading in reducing energetic and informational masking with or without perceived target-masker separation. With speech always presented from the front, the perceived location of the masker was either the same as the target's (collocated target and masker from a front loudspeaker) or near a loudspeaker located 60º to the right (masker presented through a delayed channel in front and additionally but without delay from the right loudspeaker). Due to the precedence effect, listeners perceived the masker to be located near the right loudspeaker in the separated condition Although outcome measures were percent-correct identification of sentences, enough of the psychometric functions were measured that one could derive from them changes in SRTs. In the collocated situation, LRBs of 3.5 and 8.5 dB were measured for SSN and two-talker babble maskers respectively. In the separated condition, the LRB did not change for SSN, but it dropped to 6.2 dB for the babble masker. Going from a single, collocated masker to the situation that added the masker signal in the right loudspeaker, SRTs in SSN worsened by 1 dB for both audio and AV conditions whilst SRTs in competing speech improved by 2.1 dB in AV and 4.7 dB in audio-only conditions. The authors propose that in conditions involving informational masking, over and above their benefits described in the previous section, visual cues further help disambiguate target and masking speech. The visual target information, when temporally and spatially congruent with the auditory target but incongruent with the masker, may help both focus the auditory attention onto the target and ignore the competing speech.

Helfer & Freyman's findings suggest that LRB with the talker in front may be independent of the perceived azimuthal separation of a single SSN masker within a restricted SNR range. Furthermore, provided that head orientation away from the target talker does not impair the visibility of their lips, tongue and teeth, LRB should be independent of head orientation. These expectations will be verified in Chapters 4 and 5.

1.3.2. Lip-reading benefit in CI users

CIs deliver mostly the temporal envelope of speech and not its fine structure, nor the low-frequency voice pitch usually accessible to HA users; as a result CI users are particularly susceptible to noise (see Section 1.2). Since the visual cues are unaffected by noise, CI users tend to rely more heavily on them than NH listeners. Most CI users achieve improved AV speech intelligibility over the audio-only situation (Lachs et al. 2001; Bergeson et al. 2005; Hay-McCutcheon et al. 2005; Moody-Antonio & Takayanagi 2005). Some CI users have also been shown to be superior speech readers and AV integrators (Goh et al. 2001; Clark 2003; Schorr et al. 2005; Rouger et al. 2007; Desai et al. 2008; Strelnikov et al. 2009).

Unfortunately none of the literature reports comparative audio-only and AV SRTs (in silence or in noise). Instead, studies typically measured speech intelligibility using percent-correct identification in silence of CV tokens, spondees, isolated words or sentences. We will therefore not be able to quantitatively compare our LRB outcomes (Chapters 5) with these previous studies. Nevertheless, these studies provide insight into the differences in both the lip-reading ability and the AV benefit between NH listeners and CI users as well as insight on the source of such differences.

Both Lachs et al. (2001) and Bergeson et al. (2005) studied prelingually deaf children with CIs. Lachs et al. found that children that were better at recognising isolated words in the audio modality obtained a larger AV benefit and were also better at producing speech. These correlations led the authors to propose that a common source of linguistic knowledge is used in both perception and production of speech, based on the articulatory motions of the vocal tract. Using sentences, Bergeson et al. found AV scores to be superior to scores in either modality alone. The authors suggested that lip-reading and AV speech perception reflect a common source of variance associated with the development of phonological processing skills. By testing the McGurk effect in prelingually deaf children as a function of age of implantation, Schorr et al. (2005) found that, in contrast to NH children who experience normal bimodal fusion, most CI users exhibited atypical bimodal fusion. Visual cues tended to dominate, suggesting a higher reliance on lip-reading. Bimodal fusion was however more likely when children were implanted before the age of 2.5, which implies a sensitive period of normal AV integration development.

With congenitally deaf adult CI recipients, Moody-Antonio et al. (2005) found AV intelligibility to be, as for children, equal or superior to that in either modality. While 3 of 8 participants showed an additive benefit in terms of percent-correct score, another 3 displayed a super-additive effect. Participants could significantly benefit from AV integration despite their lack of auditory experience pre-implantation. In comparing speech perception measures between elderly and younger adult CI users, Hay-Cutcheon et al. (2005) found no difference in audio-alone outcomes. However, whilst older adults were a little poorer at lip-reading, they were more efficient AV integrators. In contrast with Schorr et al. (2005) in children, Desai et al. (2008) did not succeed in finding significant differences in bimodal fusion with the McGurk effect when comparing adult CI users and CI simulation to NH listeners. A growing body of evidence (e.g. Tremblay et al. 2010; Landry et al. 2012) suggests that CI users need to be split into two groups: Proficient CI users on one hand, who tend to favour the auditory stimulus and exhibit close-to-NH bimodal fusion; non-proficient CI users on the other hand, who tend to favour the visual input and display atypical bimodal fusion. When Desai et al. made use of percent-correct identification of CV tokens by NH listeners and CI users, CI users showed significant AV benefit, but presenting vocoded speech to NH listeners failed to simulate the AV benefit to CI users. NH listeners' (non-degraded) audio-only and AV scores were unfortunately at ceiling. AV integration could therefore not be compared between listeners. The authors also found that the AV integration benefit correlated with the duration of CI experience, not that of deafness. This was echoed by Rouger et al. (2007) and Strelnikov et al. (2009) who found that the AV benefit gradually increases post-implantation over the first 2 years of an 8-year longitudinal study. Despite considerable auditory recovery over the first year, CI users were found to retain the superior lip-reading ability (over NH listeners) that they had pre-implantation. The authors felt this may have been the result of a strategy CI users develop to better cope with noisy situations. When the speech presented to the NH listeners was degraded by noise or vocoding in such a way that their audio-only performance equalled that of CI users, CI users exhibited an AV speech intelligibility superior to that of NH listeners. This was not only due to superior lip-reading, but also to superior AV integration because the CI users' AV gain was super-additive and nearly twice as high as that of NH listeners. It is plausible however, that as for bimodal fusion, should the CI users be split into proficient and non-proficient groups, proficient CI users may exhibit more 'normal' AV integration. Also, had NH listeners had time to accommodate to distorted auditory input, they may well have exhibited similar AV integration capability as that of CI users.

Caution is required when considering the degree of integration in bimodal speech recognition. Indeed, the additive or super-additive nature of integration should not be assessed directly from percent-correct performance. When considering the role of redundant (reinforcing) or complementary contextual information in speech intelligibility, Boothroyd & Nittrouer (1988; Nittrouer & Boothroyd 1990) showed how error rates, rather than percent-correct scores should be compared. The addition of visual information should parallel this conflux of information sources. Thus, with A, V and AV denoting audio-only, visual-only and AV percent-correct scores, the audio and visual contributions to speech recognition are expressed as $(AV - V)/(100 - V)$ and $(AV - A)/(100 - A)$. These ratios establish by what factor the percentage of error rate is reduced by addition of one modality to the other. Let us now denote P_X as the probability of recognising the target in the X modality. Without interaction between modalities, the probability of failure to recognise the target should be multiplicative, such that $(1 - P_{AV}) = (1 - P_A) * (1 - P_V)$. Thus, if P_{AV} exceeds $P_A + P_V - P_A * P_V$, a reinforcing interaction occurs between modalities. Correspondingly, if AV exceeds $A + V - A * V$, one has detected the super-additive nature of the AV integration. Conversely, if $AV = A + V - A * V$, integration is simply additive.

Most of the studies reviewed here clearly indicate that the above conversion from percent-correct measures to error rates was operated when comparing the degree of AV interaction that NH listeners or CI users exhibit or when monitoring how integration evolved with time post-implantation or with age. However, it is apparent throughout the literature reviewed that many studies suffer from the ceiling effects that percent-correct identification measures can bring. This is often unavoidable when the same tests are used over a large dynamic range of intelligibility. The workarounds that some authors resorted to were questionable as they made comparisons unreliable and some conclusions debatable. These problems highlight the relevance of the use of SRT measures in noise, since by definition they cannot suffer from ceiling effects. All measurements reported in our experimental chapters will make use of such measures. Furthermore, the stimuli used will need to be sufficiently complex that they cannot be fully understood through lip-reading alone, even by our best lip-readers, failing which measures of SRT would diverge towards infinitely low values. For that reason, a closed set of stimuli is not desirable and our SRT measurements in Chapters 3 to 5 will make use of open sets of sentences.

1.4. Free-head orientation, behaviour & strategies

Many have studied head movements in sound localisation experiments (e.g. Young 1931; Wallach 1940; Thurlow et al. 1967; Wightman & Kistler 1999). Infants are known to exhibit reflexive head turns in response to sounds (e.g. Muir & Field 1979) and a typical conversational turn-taking behaviour when listening is to face our interlocutor (Kendon 1967). The first study of the head movements of listeners attending to speech in noise was reported by Hirsh (1950). Although the author showed that with the head free to move, listeners could reap some HOB, the magnitudes of head movements were not measured or related to the HOB obtained. Hirsh was mostly interested in demonstrating that over and above speech-facing SRM, localisation of azimuthally separated target and masker could lead listeners to reap an additional HOB.

Brimijoin et al. (2012) may have been the first to study head movements and head orientation strategies in a speech listening task with spatially separated noise. The authors set out to systematically evaluate the role of head orientation in speech-listening strategies in noise. The authors' hypotheses were that the listener would aim for either maximum target level or maximum SNR at their better ear. A short adaptive speech-in-noise task with four reversals was employed to reach a SNR close to their SRT_{50}, i.e. one that the authors reasoned would promote head movements. This was repeated for a variety of target orientations 360° around the listener and for a range of masker separations. As the authors had failed to elicit spontaneous head movements in normally hearing listeners (pers. comm), they chose to test asymmetrically hearing-impaired listeners (> 16 dB asymmetry), reasoning that the propensity for making use of head turns in poor SNR conditions would be strongest for such listeners. An infrared motion tracking device attached to the listeners' heads enabled recording of head movements. The listeners attended to 2-3 s short sentences (Adaptive Sentence List corpus from MacLeod & Summerfield 1987) in SSN, whilst sat in the centre of a circular array of loudspeakers. Selecting head orientations adopted at the best trial point of each adaptive track (lowest SNR with successful sentence recognition), the median of the (near-Gaussian) distribution of azimuthal orientations across all trials was 50° away from speech-facing, such that the better ear was brought closer to the target. This was independent of masker separation and was close to the 60° head orientation that provided maximum target level. Although in 2 of 5 spatial configurations the head orientation for maximum target level and SNR nearly coincided, in the remainder they differed by as much as 100°. Brimijoin et al. concluded that the natural head orientation strategy when attending to speech in noise is typically to orient one's head so as to maximise target level. Although a suboptimal strategy for better ear listening, the authors argue that it is simpler to implement, especially when localisation of the noise is difficult or in complex listening situations where interferers are distributed around the azimuthal plane.

Brimijoin et al.'s study presents several issues. First, there is no equivalent study in NH listeners that can be used as a benchmark. Second, the authors claim (even in their title) that the behaviour investigated is undirected and therefore indicative of natural strategies. Although the outcomes may incidentally reflect natural tendencies, the undirected nature of the experiment is debatable. Not only are the listeners wearing a tracking device that could lead them to think that the researchers are interested in head movements, but also and by the authors' own admission, 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 conceived as an instruction or an implied direction. In contrast with Brimijoin et al., we will endeavour in our experiments to not do or say anything to participants that could give them any clue whatsoever that head orientation is a key focus of the experiment. Hence covert overhead video recording and subsequent head orientation encoding was adopted (see all free-head experiments in Chapters 3 to 5). Finally, Brimijoin et al. combined in their experimental protocol the head-orientation behavioural measure with a partial adaptive track, aiming at reaching what the authors assumed to be adequate SNRs. Although the authors clearly point out that their intention was never to measure accurate SRTs, we feel a clearer picture may be obtained by separating objective measures of SRT from head-orientation behavioural measures or subjective measures of SRM. This forms the basis of our experimental approach (see experimental Chapters 3 to 5).

The only investigation of head-orientation response to target speech presentation in CI users was reported by Van Hoesel (2015). Van Hoesel devised a novel spatial audio-visual test paradigm to assess bilateral CI users for their speech intelligibility in free-head dynamic and complex listening situations. Following an audio cue for the target position, the target speech was presented in audio or AV modality from 1 of 4 random locations spanning 180º in the frontal arc. Distracting talkers were presented visually only from the remaining 3 azimuths and 8 audio-only speech distractors were distributed 360º around the listener to form a masking sound field mimicking that experienced in a large cafeteria. In binaural conditions, inclusion of visual cues led to a 5 dB LRB. Binaural AV SRTs were 5 to 15 dB superior to those obtained with the better ear alone whilst monaural SRTs did not show any LRB, probably because visual cues were not being seen. It seems likely that monaural AV testing of bilateral CI users will have totally removed their ability to make sense of the complex listening situation and localise the audio cue or the target speech (via audition, vision of both) over the short time-span of a target sentence presentation. This would have dramatically reduced their chances of lip-reading from the correct video monitor because they did not know where to look. Hence the 5-15 dB binaural AV benefit observed is presumably made up to 5 dB LRB and 0 to 10 dB HOB. Aiming at mimicking highly dynamic and complex listening situations, this study lacks a baseline from the literature that separately demonstrates LRB and HOB for CI users in a simpler speech-in-noise situation.

To this date, we believe that no investigation of spontaneous NH head-orientation behaviour when attending to audio-only speech in noise (let alone AV speech) has ever been reported. This, in tandem with a validation of speech-facing SRM and HOB model predictions (Jelfs et al. 2011) has to be our starting point (Chapters 3). Next, one must investigate the impact of visual cues and lip-reading on the above, since bimodal perception of speech is present is most social situations (Chapters 4). Having measured a NH baseline, one can then compare objective and subjective measures between CI users and NH listeners and, should CI users make little use of head orientation in a free-head task, demonstrate how a simple instruction can help them reap a HOB (Chapters 5). The translational application of our findings will naturally ensue.

Before we proceed to the experimental chapters, given that our experimental approach is informed by the Jelfs et al. model of SRM, Chapter 2 will give the reader a brief introduction to models of speech intelligibility in noise before we delve into a detailed description of the Jelfs et al model.

2. Models of speech intelligibility in noise

Several models have been developed that consider speech intelligibility from different perspectives. Each angle of approach to the subject relies on a series of assumptions that form the basis of the model. Such assumptions may be empirically or theoretically based. The processes that the model employs may be inspired by neurophysiological evidence. Alternatively, such processes may be based on the postulates required to define a “black-box” type of model, one that takes no account of the physiology of the auditory system and does not concern itself with the intricacies of sound processing and decision making by the nervous system. In this chapter, we will first describe the monaural models that can be used to predict speech intelligibility in noise and reverberation, as perceived by a single ear. We will then consider models of BU, those that consider how the existence of two ears enables the auditory brain to analyse interaural relationships and release speech from masking beyond better-ear listening. Finally, we will examine models that combine BE and BU to predict SRM and consider how the Jelfs et al.(2011) model lends itself to predicting SRM for CI users.

2.1. Monaural models

2.1.1. The articulation index

Articulation is a measure of the proportion of phonemes uttered by a speaker that are correctly interpreted by a listener. It is linked to the intelligibility of speech via empirically-derived psychometric functions that depend on the speech material used. The articulation index (AI) aims to predict articulation (and therefore speech intelligibility) as a function of the *sensation* SNR received by the listener. It takes no account of temporal variations. A simple SNR is the ratio of speech to noise levels (with levels defined from a fixed sound-pressure-level reference). A sensation level is defined as a sound level referenced to the listener’s absolute threshold for that sound. By sensation SNR, we refer to the ratio of sensation levels for speech and noise (each referenced differently, by definition), such that hearing loss, for instance, may be taken into account through elevated audibility thresholds. The calculation of the AI also takes into account the various ways that noise can mask speech.

Most of the ground work in building the AI model was formalised in the first half of the 20th century by Bell Labs researchers, and therefore driven by telephone systems transmission characteristics and transmission service performance (Martin 1931; Munson

1945; French & Steinberg 1947). The articulation material developed up to the early 40's was made to be representative of typical speech, was used for the characterisation of transmission instruments. The drop in articulation due to distortions introduced by transmission equipment was measured and the quantitative transmission data thereby acquired was used in the engineering of the telephone plant (Martin 1931). The considerations relevant to telephonic speech transmission to a listener were considered by French & Steinberg (1947) and formed a basis for the formulation and standardisation of the AI model and method.

By expressing intelligibility relationships in terms of sensation SNR, and on the basis of standard articulation material and standard articulation data for young NH listeners, one can investigate how a given characteristic of the speech, of its transmission, of its perception, or of its masking by noise affects speech intelligibility.

2.1.1.1. Model assumptions

The AI model assumes that (1) the ear can be modelled as a bank of successive, non-overlapping, rectangular auditory filters spanning frequencies relevant to speech perception; (2) intelligibility of speech can be entirely accounted for by separate contributions from each band and (3) the contribution from each band is independent of the contribution of another. Thus, the AI assumes that any narrow band of speech frequencies of a given intensity contributes to the intelligibility of speech independently of the other bands with which it is associated, and that the total contribution of all bands is the sum of their individual contributions.

The assumption of cross-band independence is not strictly true when neighbouring frequency regions carry intense speech and cause masking of a quieter band but the impact of cross-band masking was considered negligible in the original formulation by French & Steinberg (1947). The vibration mechanics of the basilar membrane in the cochlea are such that when the membrane is excited by a given tone, its vibration amplitude is maximal at a particular *place* along the membrane and decays rapidly towards more apical, lower-frequency places, but much more slowly towards more basal, higher-frequency places (e.g. Moore 2012). As a result, although a given narrow band of noise will mask most effectively the matching band of speech, it will also mask speech upward in frequency, i.e. in more basal places along the basilar membrane. This gives rise to the *upward spread of masking*, which was identified long ago by Weger & Lane (1924) and must be taken into consideration when working out the AI. A remote downward masking phenomenon identified by Bilger & Hirsh (1956) is also considered, although often negligible for broad-spectrum noise.

2.1.1.2. Methodology

The calculation of the AI consists of determining which bands are masked by a background noise, knowing the speech and the noise spectrum levels in each band and how masking extends upwards and downward in frequency. Three similar methods were developed, one originally based on 20 bands of equal contribution to the AI (French & Steinberg 1947), the other two based on splitting a logarithmic frequency scale into bands of equally spaced centre frequencies: octave bands (OBs) and third-octave bands (⅓OB). The OB and the ⅓OB versions of the AI method are simple modifications of the 20-band method. These versions were included in the ANSI standard (1969) because filters suitable for the 20-band method were not available in practice, whereas OB and ⅓OB filters were easily accessible. With the 20-band method, the 20 *equally important* bands span 200-6100 Hz in a manner that the centre frequencies of the bands are not equally spaced along a logarithmic scale, in contrast with OBs or ⅓OBs. The frequency importance to speech intelligibility depends on frequency. The function peaks in the 1600-2000 Hz range (the *cross-over* or *importance frequency* in quiet, see Pollack 1948; Dyer 1962; Webster & Klumpp 1963) on a logarithmic scale.

In its simplest form, and based on the original assumptions, the AI was expressed by French & Steinberg (1947) as the sum of the contributions of 20 spectral bands of equal and maximum contribution $\Delta A_{max} = 5\%$, as follows:

$$(1) \qquad A = \sum_{n=1}^{20}(\mathit{\Delta} A_n) = \sum_{n=1}^{20}(W_n \,.\mathit{\Delta} A_{max}) \,,$$

with a contribution ΔA_n to the AI of the n^th^ frequency increment Δf_n, W_n being the fractional amount of the maximum ΔA_m available that depends on the effective sensation level for band *n* (i.e. how much above threshold the signal is). The expression of AI could then be simplified from equation *(1)* as:

$$(2) \qquad A = \frac{\sum(W_n)}{20}$$

In effect, W is a container that expresses the result of all sources of degradation of speech intelligibility along a distorted frequency scale that satisfies [$\Delta A_{max} = 5\%$] for each frequency band.

Specifically applied to the intelligibility of speech in continuous noise, with the total dynamic range of speech levels between minima and peaks estimated to span 30 dB (Beranek 1947), the effect of a continuous noise on the AI contribution within each band could be expressed as:

$$(3) \quad W_n = \frac{(level\ of\ speech\ peaks) - (level\ of\ noise)}{30} \quad \& \quad 0 \leq W_n \leq 1$$

Given a full dynamic range of speech of 30 dB, a requirement for the AI to reach unity without noise is that the peak value of the speech signal in each band be at least 30 dB above the threshold of audibility of continuous-spectrum sounds. In other words the speech should be clearly audible. In order to take into account the intelligibility degradation by noise and audibility thresholds, the procedure to be followed essentially inflates the noise spectrum to reflect the upward and downward spreads of masking. The difference, at the central frequency of each band, between the speech peak level and the higher of (1) the inflated noise level or (2) the threshold level for the audibility of a broadband noise, is then added up along the 20 bands. One has to cap each contribution to 30 dB (the full dynamic range of speech) and null it when the speech lies below the noise level and/or audibility threshold. Since perfect intelligibility (A = 1, with speech entirely above audibility threshold and without masking) would add up to 20 x 30 = 600, the sum above is divided by 600 to compute the AI for the particular speech and noise spectrum levels and level of hearing impairment considered.

Tables or weighting factors for the octave and ⅓-octave bands are used to operate the conversion from the 20-band method. The rest of the manipulation remains the same. Kryter (1962a; 1962b) argued that the OB methods can lack resolution and result in inaccurate AI estimates because of the relative insensitivity of a small number of OBs to sharp changes in the spectrum shape of the speech signal or of a masking noise. It must therefore be used with some caution. Kryter validated the AI by providing an early demonstration that the 20-band and ⅓OB methods proved accurate enough for most applications (see Section 2.1.1.4).

Fletcher & Galt's (1950) general method had provided an earlier framework for the estimation of the AI as a function of a wide range of factors. Their method was a means to approximate the effect, amongst others, of a range of frequency or amplitude distortions (from e.g. a telephone transmission line, reverberation, hearing impairment or saturation). Over the following decades, theoretical developments as well as the deepened understanding of physiological mechanisms underlying speech perception in noise led to Fletcher and Galt's general method not being retained. However, one concept they introduced, that is most relevant to this thesis, is that of proficiency of a listener, reflected by an overall-AI multiplicative *proficiency factor* P, between zero and unity. Indeed, the proficiency factor can be used to reflect hearing loss as we will see later (see Section

2.1.2.1). Another is the concept of *effective AI*, an AI that reflects the boosting of speech intelligibility by the presence of visual cues and by our lip-reading ability.

2.1.1.3. The relationship between AI and speech intelligibility

Once the AI has been calculated with all sources of speech intelligibility degradation compensated for (where possible), speech intelligibility is derived from established curves that link the AI to it. Fletcher & Galt (1950) specifically focussed on the relationship between the AI and intelligibility of units of speech (vowels, consonants, syllables, words) or sentences. The authors systematically analysed and compared data sets acquired with four different articulation tests designed over the previous thirty years and that measured articulation of speech for various sequences of consonants (C) and vowels (V) (CVC, CV or VC syllables or a mixture of the different types). Such tests typically employed a carrier sentence aiming at focusing the listener's attention and spoken before the syllable was called (e.g. "You may perceive…" or "I was about to say…" before calling [*na'v*] or [*bĕk*]). The authors empirically linked syllable articulation to sentence articulation. The relationship between AI and intelligibility resembles a sigmoid since it is a psychometric function of intelligibility versus AI. Kryter (1962a; 1962b) explains how, the greater the constraints inherent to the material (in its range, context or grammatical structure), the greater the intelligibility for a given AI. For instance, intelligibility reaches 50% at AI = 0.38 for nonsensical syllables, at AI = 0.34 for a set of 1000 PB words, at AI = 0.19 for a set of 256 PB words, at AI = 0.15 for sentences and at AI = 0.125 for a closed set of 32 PB words. These relationships illustrate the range of maximum AIs that one would expect for SRTs measured with a given set of material, and in other words, how material characteristics can affect measures of SRT.

2.1.1.4. Early validation of the AI

Kryter (1962a; 1962b) reported that the AI could provide accurate predictions of speech intelligibility with a steady state noise and with a range of distortions of the speech. The author lists the types of speech or noise distortions that he claimed could be compensated for in the AI calculation. With *noise amplitude fluctuations*, we know from Miller (1947) that speech intelligibility improves (see Chapter 1). Kryter argued that the nature of the fluctuation determines whether the AI calculation can be adjusted in a simple manner. With a definite on-off duty cycle, he claimed that the AI can be corrected. However, his approach was somewhat crude since the AI reduction presented as a function of duty cycle was based on empirical data from Miller (1947) and on the basic assumption that the effect of duty cycle is simply additive with the effect of modulation rate measured by Miller & Licklider (1950). Kryter's approach was not as theoretically motivated as later models would be (see e.g. Rhebergen & Versfeld 2005 in Section 2.1.2.3). Kryter acknowledged that corrections were not straightforward when fluctuations are more complex, e.g. irregular in time and amplitude. Kryter proposed that *Frequency distortions*, i.e. varying gain as a function of frequency (often encountered in telecommunication devices) can be compensated for, providing their emphasis is restricted to certain regions of the speech or noise spectra and providing they do not contain multiple peaks and valleys with slopes exceeding 18 dB/octave on average. Again here, no theoretical basis for this claim was presented as the proposed rules were purely empirical. Kryter argued that *amplitude distortions* such as peak clipping and also the effects of *reverberation* can be taken into account. For reverberation, he empirically linked the amount of intelligibility degradation to the reverberation time (RT_{60}) in a room. It should be noted, however, that RT_{60} is not a reliable indicator of acoustic quality from the standpoint of speech intelligibility. For instance, a broadband RT_{60} value can conceal variations across frequency, which would influence intelligibility. A later, theoretically motivated alternative to the AI, the speech transmission index (see Section 2.1.2.2) would better account for the effects of reverberation. *Vocal effort*, resulting in weak or intense speech levels, Kryter claimed can be factored in, providing it does not fluctuate. Even *visual cues* that improve intelligibility through lip-reading or speech-reading (see Chapter 1) could be taken into account in the calculation of an *effective AI*. Again here, Kryter presented a conversion graph based on empirical data.

2.1.2. Extensions of the AI model and alternative approaches

The AI was renamed the speech intelligibility index (SII) in the early 90's. Following its introduction, a number of research groups have had mixed success in validating or extending the AI/SII model for the intelligibility of speech in more complex, but ecologically relevant situations (e.g. Dubno et al. 1984; Grant & Braida 1991; Ching et al. 1998; Rhebergen & Versfeld 2005). Where required, an alternative to the AI approach was formulated (e.g. Steeneken & Houtgast 1980; Elhilali et al. 2003; Jørgensen & Dau 2011). One can split the sources of speech intelligibility degradation into two types: (1) those affecting the signals only in the frequency domain (2) those affecting the signals in the time domain. The AI method is appropriate for distortion in the frequency domain since it is based on the SNR at the ear. However, it is typically not adequate for distortions in the time domain. Among time-independent impediments other than masking by steady noise, most relevant to this thesis is the impact of hearing loss and age on the AI accuracy. Audio-visual speech intelligibility is known to be superior to that of audio-only listening (see Chapter 1). The related improvement can be modelled with some success by adjusting the AI. Time dependent impediments include, fluctuating noise or competing voices and reverberation of the speech signal. Both require different modelling approaches, but our experimental endeavours will be restricted to examining the effects of steady state noise and the effects of reverberation of that noise.

2.1.2.1. AI predictions for hearing loss and age

Fletcher & Galt (1950) used a proficiency factor to characterize the enunciation of the talker and the experience of the listener with that talker. This scaling factor, ranging from zero to unity, was to be applied after the summation of AI contributions by individual bands and hence affected all bands equally. Fletcher (1952) proposed that the proficiency factor could be used to describe the effects of hearing loss other than reduced audibility. Dugal et al. (1980) adopted the approach, and reported that the effects of frequency-gain characteristics and signal levels were well predicted for a group of six hearing-impaired listeners when the AIs were rescaled by individually derived proficiency factors.

Dubno et al. (1984) investigated how adequately the AI could reflect the combined effects of age and mild HL. The authors compared AI predictions to empirical data acquired with high and low-predictability sentences from the Speech Perception in Noise (SPIN) corpus (Kalikow et al. 1977). The SPIN sentences were presented in quiet or in a multi-talker babble resembling SSN. The ⅓OB method was used and hearing impairment reflected by ⅓OB auditory thresholds interpolated from audiometric data. No significant effect on AI predictions was found with age or hearing loss in quiet, which reflected empirical results for all four factor combinations (younger/older adults * normal/impaired hearing). Differences in performance due to HL were adequately predicted in noise. However, higher AIs were predicted for the older groups relative to their younger counterparts. This was a surprising consequence of combining pure-tone thresholds with the relative levels of speech and noise required for 50% performance. As expected, empirical results exhibited lower performance with age and the age effect was found to be independent of HL. These findings highlighted a limitation of the AI as it failed to reflect age-related speech intelligibility deterioration. Age must therefore be compensated for in the AI calculation.

Ching et al. (1998) studied predictions in HI listeners and the limited role of high frequency amplification. For mild to moderate HL, the AI was found to reflect performance reasonably well. However, in cases where regions of severe to profound HL existed, the AI greatly overestimated intelligibility at high sensation levels and could underestimate performance at low sensation levels. Audibility could not explain speech recognition in many cases. The proficiency factor needed to be made smaller at high levels than at lower levels. Furthermore, the measured scores for broadband stimuli exceeded predictions based on the summed contributions of OBs spanning the same bandwidth. This indicated that the AI assumption of independent band contributions was often violated. The best attempts to make the model fit the data combined the effect of standard level distortion factor (seen at high levels in NH listeners) with an individually derived, frequency-dependent proficiency factor. For regions requiring high amplification according to the AI, the contribution of audibility was much reduced and sometimes non-existent. Further increases in audibility could even worsen speech intelligibility. This finding made the authors debate the validity of amplification of severe HL regions. Moore & Glasberg (1998) proposed an alternative to the use of the AI for prescribing insertion gain from audiometric thresholds. The *Cambridge formula* aims to restore *loudness* to levels similar to those evoked in NH listeners by 65 dB speech and leads to higher adjusted AIs than the original AI method. When a high frequency hearing loss is compensated for by a hearing aid, care must also be taken to take into account the possible complete loss of inner hair cells in regions of the cochlea (the dead regions). Vickers et al. (2001), Moore (2002) and Baer et al. (2002) indeed showed how the incremental benefit of amplifying frequencies well above the estimated edge frequency of a dead region is overestimated by the AI. They highlighted the importance of diagnosing dead regions and show that a modified version of the AI can help prescribe improved hearing aid insertion gain.

2.1.2.2. The speech transmission index approach

The speech transmission index (STI) does not predict speech intelligibility from the SNR as the AI does. It is instead based on the idea that speech intelligibility is correlated with the degree to which the speech amplitude modulation (within each frequency band) is preserved by the system that transmits the signal. As for the AI, such a system could be a telephone line or two people conversing in a room.

(Steeneken & Houtgast 1980; 1983) introduced the STI as an AI alternative that accounts for disturbances in the time domain and for non-linear distortions. Within each OB, a channel (e.g., a recording or transmission medium, a room or a vocoder) is probed with a test signal, so as to establish its transmission characteristics. The modulation index of the test signal fed through the channel is modified by parallel transmission of noise. A matrix of modified modulation indices is generated that is converted into a transmission index (TI). Combining the TIs over seven OBs with weighting factors comparable to those used in the AI calculation leads to the STI. Since distortions in the time domain affect the envelope of the signal, for each OB the modulation index has to be evaluated as a function of modulation frequency. Each OB leads to a number of modulation-frequency-specific TI values. The contribution of each OB to the STI is called the modulation transfer function (MTF), a measure of the fidelity with which a channel transmits different modulation frequencies. The MTF averages the TI values over a range of modulation frequencies chosen to reflect the characteristics of speech envelope spectra (the authors prescribed ⅓OBs in the 0.63-12.5 Hz range).

The STI only reflects the effect of the strength of signal modulations. Reverberation impacts both speech and noise. The MTF and STI can be used to model it, as was demonstrated by Houtgast & Steeneken (1985) for reverberant auditoria that cause temporal smearing of the signal. Temporally-varying sources of intelligibility degradation that are particularly ecologically relevant are fluctuating noises and interfering speech. With the STI approach, the effect of noise is not accounted for through a measure of SNR, but through its effect on the modulation transfer function. Although successful for steady noise conditions, the STI is consequently completely unable to deal with modulated noise, because the modulation of the noise contributes to the apparent MTF of the signal. An alternative calculation of the STI is therefore required. In the case of noisy speech processed by spectral subtraction, since the STI makes no distinction between speech and noise fluctuations, it cannot account for the masking release effect.

2.1.2.3. STI, AI and alternative model variations for fluctuating noise

Ludvigsen et al. (1990) introduced a novel method for the calculation of the STI that aimed at avoiding artefacts introduced by the STI when the signal was processed by non-linear devices such as hearing aids. With speech used as the input signal, the new method computes the STI fluctuations (over 23 bands) from the signal intensity envelopes of the noise-free speech signal and of the processed, noisy signal. Thus, the new method takes into account both system non-linearity and masker temporal fluctuations. The STI calculated from the original Steeneken & Houtgast's (1983) method for a linear system and stationary noise was exactly replicated. Moreover, the new method was satisfactorily validated against a number of situations with non-linearly processed speech or fluctuating noise. Festen & Plomp (1990) studied the effects of fluctuating noise and interfering speech on the SRTs of NH and HI listeners. When interfering noise was modulated with the wideband envelope of a second voice, the Ludvigsen et al. model predicted reasonably well the shift in SRTs due to dip-listening. The threshold shift for an interfering voice was however underestimated, which is understandable since the calculations did not take into account linguistic cues or the exploitation of differences in F0 (see Chapter 1).

An extension of the STI inspired by neurophysiological recordings was presented by Elhilali et al. (2003) and named the spectro-temporal modulation index (STMI). The STMI analyses the effect of noise, reverberation and other distortions on the spectro-temporal modulations present in speech and characterises the ability of a channel to transmit such modulations. The STMI includes a two-dimensional modulation filter bank that analyses the spectral modulations of the speech signal in addition to the temporal modulations. It does not assume channel independence. The STMI reflects the deterioration in the spectro-temporal modulation content of speech due to any added noise or reverberation. A second method replaces speech samples with spectro-temporal ripples, i.e. amplitude-modulated sinewaves that oscillate concurrently along the frequency and time axes. The ripple-based and speech-based STMI were shown to be consistent with the STI in conditions with additive noise and reverberation. Furthermore, both STMI versions could account for the effects of phase jitter and phase shifts, two distortions to which the STI is insensitive. The STMI account of phase distortions is effected by an analysis of the modulations across the frequency axis that was not included in the STI. However, the STMI presents the same limitations as the STI for noisy speech processed by spectral subtraction.

Rhebergen & Versfeld (2005) presented a variation of the AI model designed for fluctuating interferers and named it the extended SII model (ESII). The approach consists in partitioning both target speech and masking noise into small time frames within which a conventional AI can be derived. The AI is subsequently averaged over time. The time frames were carefully chosen by the authors to be compatible with the frequency-dependent temporal resolution of the auditory system and avoid inaccurate estimations in the low-frequency bands resulting from excessively short time frames. The best compromise was found with time frames varying from 35 ms at 150 Hz down to 9.4 ms at 8000 Hz. This approach had some success with fluctuating noise and 4-and-above-talker babble, but was again limited for voice interferers (one or two) and with sine-modulated noise for some modulation frequencies similar or larger to those present in speech. Taking into account forward masking in the ESII model was shown to improve prediction accuracy (Rhebergen et al. 2006). Further conditions employing a range of real-life noises chosen on the basis of their spectro-temporal characteristics also produced ESII predictions superior to AI-model predictions although some still substantially deviated from SRT data (Rhebergen et al. 2008).

In contrast with the STI, which considers the reduction in the envelope energy of speech, Jørgensen & Dau (2011) demonstrated that a metric based on the SNR in the envelope domain (SNR_{env}) was both highly correlated to the intelligibility of noisy speech and consistent with the STI in stationary noise and reverberant conditions. The SNR_{env} metric is calculated within the framework of the speech-based envelope power spectrum model (sEPSM). The main difference with the STI is the explicit consideration of the envelope noise floor, which is increased after spectral subtraction and was proposed by Dubbelboer & Houtgast (2008) as a key factor in intelligibility reduction by noise. Since the SNR_{env} metric is calculated from a long-term integration of the stimuli, the sEPSM fails in conditions with fluctuating interferers. Although the noise modulation is contained in the SNR_{env} calculation, its effect on intelligibility is not reflected correctly by the model. The authors' illustration of that limitation is that if the noise was, for instance, amplitude modulated at a rate of 4 Hz, this would increase the noise envelope power and lead to a lower SNR_{env} at 4 Hz. The contribution to the overall SNR_{env} from the 4 Hz modulation filter would be reduced compared to the stationary-noise case, and the model would predict decreased speech intelligibility, contrary to the experimental data. Thus, dip-listening cannot be accounted for by the sEPSM model in its 2011 form.

By using a time-partitioning approach similar to Rhebergen & Versfeld (2005), Jørgensen & Dau (2013) extended the sEPSM model for fluctuating interferers. The extended model accurately predicted intelligibility for speech in fluctuating noise as well as noisy speech processed by spectral subtraction (where the STMI fails) but failed to reflect the effect of phase-jitter distortion that the STMI could predict.

Overall, it is clear that no single solution to all interference and distortion cases exists, but most ecologically relevant situations can be addressed in one way or another.

2.1.2.4. Predicting audio-visual speech intelligibility

The ANSI AI & SII standards (1969; 1997) provides a framework for prediction of an audio-visual, *effective* AI. The relationship between the AI and the effective AI was derived from Sumby & Pollack (1954). Grant & Braida (1991) evaluated the ANSI AI correction procedure for the addition of visual cues (ANSI 1969 - R1986). The authors tested NH listeners attending to audio-only, AV and visual-only presentation, in noise and with a variety of band-pass filters, of sentences from the Institute of Electrical and Electronics Engineers (IEEE) "Harvard" corpus (Rothauser et al. 1969). Since the AI and SII standards were established from averaging outcomes over several listeners and talkers, deviations were inevitably found between data and predicted AV scores for individual subjects or talkers. For filtered speech with audio AI below or above 0.25, the authors found that the effective AI tended to underestimate and overestimate outcomes respectively. This may plausibly have been due, in part, to the difference in material used between this study and the basis for the effective AI. For instance the contextual cues inherent to the IEEE sentence corpus may have been more effective at helping the listener correctly identify key words in difficult listening conditions when reinforced by visual cues. Conversely, as intelligibility increases, the contribution of contextual cues may well have dropped. As the authors argue, the absolute contribution of lip-reading to intelligibility is greatest when the auditory channel is greatly degraded. It is unclear why these effects were not apparent from Sumby & Pollack's data when the effective AI was derived from them. The authors also highlight the ceiling effects that skewed the data for the higher AIs, as often seen in percent correct AV speech intelligibility measurements (see Section 1.3.2). Grant & Braida point out that speech-reading can be aided in cases of severe to profound HI by a variety of stripped-down auditory signals. For heavily filtered speech such as speech-frequency-modulated sine waves matched to F0 or formant frequencies or for amplitude-modulated sine waves matched to the speech envelope, an estimation of the AI is, however, difficult. Stripped-down auditory signals require extensive training before they can be helpful. For a time-efficient estimation of their effectiveness, Grant & Braida propose that more analytic test material (consonants and vowel segments) could help evaluate reception of phonetic cues and that, by combining such tests with supra-segmental tests that could additionally evaluate reception of stress and intonation patterns, one could estimate post-training speech intelligibility of newly-developed stripped-down auditory signals. In that sense, the data presented by the authors could be used for such estimations.

Grant & Walden (1996) focused on the adequacy of effective AI predictions for consonant recognition. They hypothesised that, given lower frequency bands carry visual cues that are complementary to acoustic cues (e.g. voicing and manner of articulation, see Section 1.3.1.3), they may provide more benefit to intelligibility than those redundant cues (e.g. place of articulation) carried by high frequency bands, at least for consonant recognition. Grant & Walden found that, while the ANSI standard assumes that filter conditions resulting in the same AI would produce the same effective AI, low frequency bands actually tended to contribute more to the effective AI than to the AI. The difference between effective AI and AI was also found to be negatively correlated with the degree of AV redundancy. In summary, it seems that more complete models of AV integration such as those briefly discussed in Section 1.3.1.3 are required to more accurately account for the impact of the complex relationships and interactions between auditory and visual cues on speech intelligibility.

2.1.2.5. Predicting CI users' speech intelligibility

Whilst the AI has been extensively used to model speech intelligibility for NH listeners and mildly to severely hearing-impaired listeners with hearing aids, much less has been published regarding its application to CI users. Because electrical stimulation provides speech information over a number of electrodes in a manner very different to acoustical stimulation, the frequency importance function (that determines the weighting factors for each band contribution to the AI) needed to be assessed for CI users and compared to that of NH listeners. A first study that informs how AI or SII may need to be altered for CI users was conducted by Mehr et al. (2001). The authors' intention was foremost to validate a time-efficient method of frequency importance estimation based on the correlational method (Lutfi 1995) and proposed by Doherty & Turner (1996) and Turner et al. (1998). From data acquired across six bands and (only) six CI users, Mehr et al. concluded that CI users in their study yielded very different weighting functions compared to the NH (almost flat) baseline. Other than the most apical and basal bands (300-486 and 3384-5500 Hz) generally bearing a lower relative weight for CI users, the large cross-participant variability over the four intermediate bands did not allow the authors to draw any firm conclusions. Henry et al. (2000) demonstrated with 15 CI users that the *reduction factor* (the proficiency factor applied at the frequency-band level in the AI calculation to reflect the poorer ability of CI users to perceive speech compared to NH listeners) was around 0.45 and frequency-independent in the 170-2680 Hz range but was significantly higher (0.65) in the 2680-5744 Hz band. The authors also showed that the assumption of independent contribution of bands to the AI was violated for the poorer proficiency CI users. As was found by Ching et al. (1998) for HI individuals, summing individual band contributions fell short of adding up to the measured AI.

Because of the high variability in proficiency, nerve survival, implant insertion depth and electrode mapping seen in CI users, and due to the resulting distortions and frequency band interdependence, no single SII standard can be generated that could be representative of all CI users. For the purpose of modelling SRM, it can be argued that the reduction factor should not have any appreciable impact on SRM since its effect will be mostly cancelled out in relative measures of SRTs (measured at 50% intelligibility). The same cannot necessarily be said about the impact of frequency bands interdependence, but since it is not well understood or could vary wildly between CI users, the modelling of SRM within this thesis will be based on NH SII weightings, as per Culling et al. (2012). More in-depth considerations are presented in Section 2.3 below.

2.2. Models of binaural unmasking

In this section we present the main two competing models of binaural unmasking and their more recent extensions. A detailed categorisation of models of BU can be found in Colburn & Durlach (1978).

2.2.1. Nomenclature for conditions used in BU measurements

BMLDs, by definition, measure the drop in (i.e. improvement of) the detection threshold of a signal (e.g. a tone) in noise when a change is introduced between the ears in the signal, in the noise or in both. When studying the effect of IPDs on detection thresholds, BMLDs measure the effect of introducing a phase difference between the ears in the signal, in the noise or in both. For a masking noise (N) and a signal (S) with IPD values $\varphi_n$and φ_s, respectively, the corresponding condition is denoted as $N\varphi_n S\varphi_s$, where the subscripts reflect the relevant IPDs.

A BMLD reference condition is typically diotic, i.e. with identical waveforms in both ears. It can also be monaural since the detection threshold is typically the same for monaural and diotic conditions. In the diotic condition, the absence of IPD is denoted as N_0S_0. In the case where the noise is in phase across the ears (i.e.presented diotically) and the signal is out of phase, the signal in one ear is offset by a phase of π radians compared to the other ear. Such a situation is denoted as N_0S_π and is often measured, since it typically produces the largest measurable BMLD. The reverse situation is denoted as $N_\pi S_0$.

The above conditions assume that the noise is coherent across the ears. When the noise interaural coherence (defined as the maximum of the cross-correlation function of the noise waveforms at the two ears) departs from unity, the noise subscript in the condition notation is replaced with the coherence symbol ρ, as, for instance, in the $N_\rho S_\pi$ condition. Whenever noise or signal is presented monaurally, the subscript is replaced with the letter 'M' (e.g. in N_0S_M). If the noise is uncorrelated, the noise subscript becomes 'u' (e.g. in N_uS_π).

2.2.2. Equalisation-cancellation theory

The Equalisation-cancellation (EC) theory, first introduced by Durlach (1963) and later revised by the author (Durlach 1972) is a black-box model that predicts the effect of binaural unmasking by computing predicted BMLDs. As seen earlier (see Section 1.2) and typically, BU is the result of differing signal and masker ITDs, a difference seen in practice when the signal and masker are spatially separated. The assumptions of the model are that the auditory brain can, within a number of frequency bands, (1) attenuate/amplify and delay the signal arriving at one ear so as to equalise the noise in that ear with the noise in the other ear, and (2) subtract the resulting signals so as to cancel the noise. A perfect or ideal EC process would totally reject the noise and lead to detection of a signal as though it were heard in quiet. This would cause BMLDs to reach levels never measured in practice. The model therefore needs to incorporate internal noise that renders the EC process non-ideal and its output representative of experimentally measured BMLDs. Postulates are required that describe the range of delay and attenuation transformations readily available to the brain as well as the internal noise mechanisms, so as to make the model fit experimental data with a minimum of parameters.

2.2.2.1. The idea at its core

> *"The above experiments"* (see the *cone of confusion* in Section 1.2.1) *"suggest that the brain preserves and evaluates time delays (perhaps by the mechanism of delay insertion in one or the other of the nerve paths between the ear and the brain) to achieve not only the directional localization of sound but also the observed discrimination against reverberation and background noise."…"If the brain could introduce at will a **time delay** in either of the nerve paths connecting each ear to the brain, the directional pattern of the two ears as a combination could be 'steered' so that maximum response could be 'aimed' in a given direction. Aiming the pattern could favor sounds coming from a given direction over those coming from other directions. This 'direction finder effect' could be made considerably more sensitive if the brain were able furthermore to **subtract the signal in the two ears**. For no phase delay in either channel this would produce a directional pattern consisting of a null or minimum straight ahead and by varying the amounts of time delay, this null could be pointed in different directions."*
>
> *Kock (1950)*

Kock's original idea is central to the equalisation-cancellation (EC) model developed by Durlach (1963; 1972) and described in the following section.

2.2.2.2. The EC theory ‘black-box’ model of BU

The EC model is constructed with four basic components: (1) a bank of band-pass filters, analogous to a set of critical bands; (2) an equalisation process between the ears on the masking component; (3) a masker cancellation process between the ears through simple subtraction of the result of equalisation; (4) a decision device that compares/combines the direct input from each filter and the processed input from the EC mechanism and produces a response. An illustration is provided in Figure 2.1. Durlach (1963) proposed that imperfections of the system could be accounted for by addition of a random *jitter* in the EC mechanism, a form of non-additive internal noise. The difference between the SNR at the EC stage output and the SNR at the output of a given ear/filter-bank represents the change in SNR due to the EC mechanism relative to that ear, which Durlach called the EC factor f. At this stage, the model implies great simplification by neglecting a number of sources of variability. It assumes the black box has *a priori* knowledge of the stimulus characteristics and complexity, such that each stage is optimised for its own operation and optimally presents information to the next stage. For the detection of a tone signal in a random Gaussian noise, Durlach arrived at a formulation of the EC factor that is expressed solely as a function of the centre frequency of the tone ω_0, the difference between tone and noise ITDs, the ratio of tone and noise ILDs and two model parameters, σ_ε and σ_δ, that quantify the standard deviation of the amplitude and time-delay errors produced by the internal jitter of the EC stage. In its original version, the model assumes the availability of an arbitrarily large store of delay and attenuation transformations. It further assumes that the statistical distributions of the jitter parameters are independent of the transformations required.

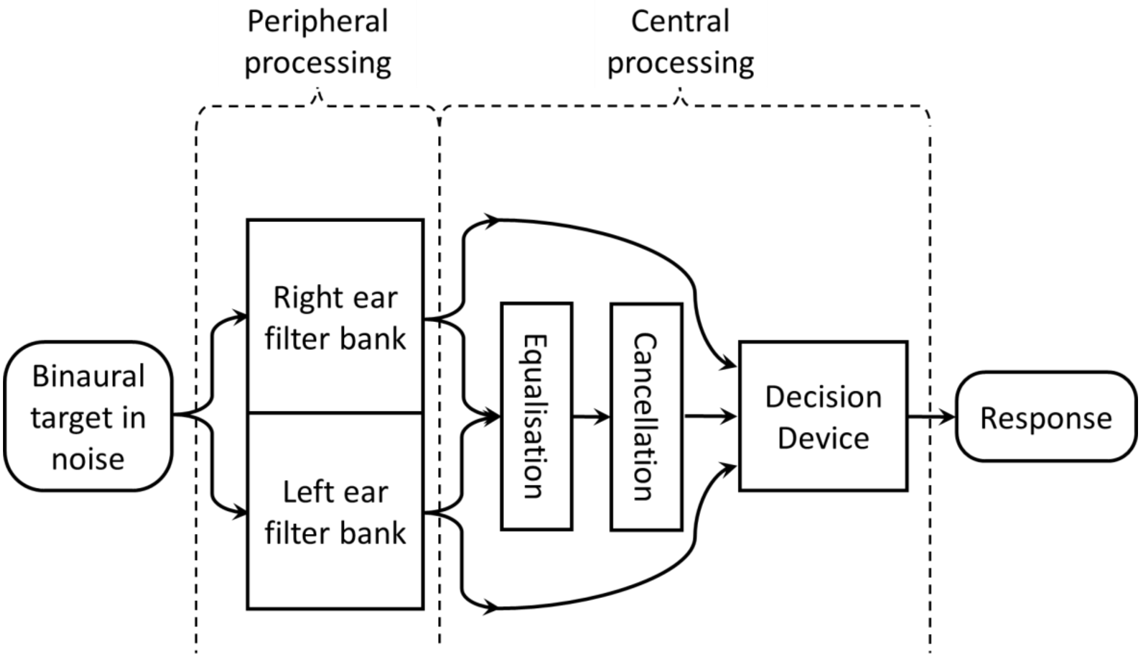


Figure 2.1: Durlach's black-box EC model schematic

Although adequate for a number of diotic-noise situations, the early model failed to correctly reflect empirical data when the masker is not the same at both ears. To address this issue, the revised model (Durlach 1972) considers that the EC stage has access to a repertoire of transformations restricted to those corresponding to ordinary experience. For instance, delays exceeding those naturally occurring as a result of the size of the head would not be normally accessible (unless unnaturally learned over a period of time). Durlach also found that it was in some cases necessary to allow the statistical distributions of the jitters to depend on the effected transformations. For a single, coherent noise and no reverberation, a BMLD (in dB) can be derived as 10Log_{10} of the ratio of the EC factors for the two conditions of interest, with the EC factors substituted with unity if lower than 1 (in other words, taken into account only when the EC process improves the SNR). With an interaural phase difference ($\varphi_s - \varphi_n$) between signal and masker at the band's central frequency ω_0, a BMLD relative to N_0S_0 can be simplified to:

$$(4) \qquad BMLD\ (\varphi_s, \varphi_n) = 10\, log_{10} \left(\frac{k - cos(\varphi_s - \varphi_n)}{k - \gamma} \right),$$

In equation *(4)*, γ reflects the envelope of the interaural cross-correlation of the noise at a given interaural delay (φ_n/ω_0). The rate reduction of γ from unity as a function of delay is greater the wider the filter bandwidth. The parameter k is defined as:

$$(5) \qquad k = (1 + \sigma_\varepsilon{}^2) exp(\omega_0{}^2 \sigma_\delta{}^2) \,,$$

and its reduction from unity expresses the extent of the internal jitter.

The expression in equations *(4)* requires the noise to be fully coherent between the ears, failing which the cancellation process would not work as well. The model can be tuned to fit psychometric data by adjusting the standard deviations of the internal noise amplitude and delay jitters (parameters σ_ε and σ_δ).

Reverberation reduces the noise interaural coherence, which is reflected by γ growing and the BMLD decreasing with increasing reverberation.

2.2.2.3. Model validation for BMLDs

Durlach (1963) examined BMLDs acquired in a variety of situations. Fitting available data for N_0S_π BMLDs relative to N_0S_0 as a function of $f_0(= \omega_0/2\pi)$, the best fit was obtained with $(\sigma_\varepsilon,\sigma_\delta) = (0.25,105\mu sec)$. Above $f_0 \approx 1.2$ kHz, however, BMLD predictions and data diverged, reaching asymptotes at the highest frequencies at 0 and 3 dB, respectively. The discrepancy between model and data beyond 1.2 kHz was best explained by the model being entirely reliant on IPDs. At higher frequencies the auditory system only encodes the envelope of the waveform. The EC model therefore needs a more realistic peripheral model rather than its phase jitter mechanism in order to adequately fit data beyond $f_0 = 1.2$ kHz.

2.2.3. Colburn’s biological-evidence-based model of BU

2.2.3.1. Model structure

Colburn built a theory of binaural interaction (Colburn 1973) by combining the modelling data from previous physiological studies in cats (Kiang 1965; 1968) with earlier mathematical modelling of the peripheral auditory system (Siebert 1968; 1970). Colburn compared the information that could be extracted from the responses of populations of auditory-nerve fibres to performance in binaural detection and interaural discrimination experiments. Colburn’s model consisted of two parts (see Figure 2.2): a model of auditory-nerve activity and a central processor that analyses and displays comparisons of nerve-firing times from ear to ear. The latter component is similar to Jeffress' (1948) place theory of sound localisation by ITDs. Colburn wrote that his model could be considered as a quantification and elaboration of Jeffress’ (Colburn & Durlach 1978). The central processor is broadly analogous to interaural cross-correlation (CC).

The first stage of the model predicts auditory-nerve patterns in man. This stage is initially designed to fit the auditory nerve firing patterns in response to 300 ms tone bursts with slow rise and decay times (50 ms). It assumes that (1) the transduction from the input sound waveforms to the firing pattern of auditory nerve fibres is probabilistic in nature, (2) individual nerve fibres are frequency-selective and have a characteristic frequency (CF) and (3) neural firing patterns are phase-locked to the stimulus. Colburn selected the non-homogeneous Poisson process to characterise the response of auditory-nerve fibres to sound because it is the simplest stochastic process that can realistically be applied to model firing patterns. Each fibre is characterised by a rate function (dependent on the spectro-temporal characteristics of the stimulus and on the CF of the fibre) that describes the instantaneous firing rate assumed to be produced by that fibre.

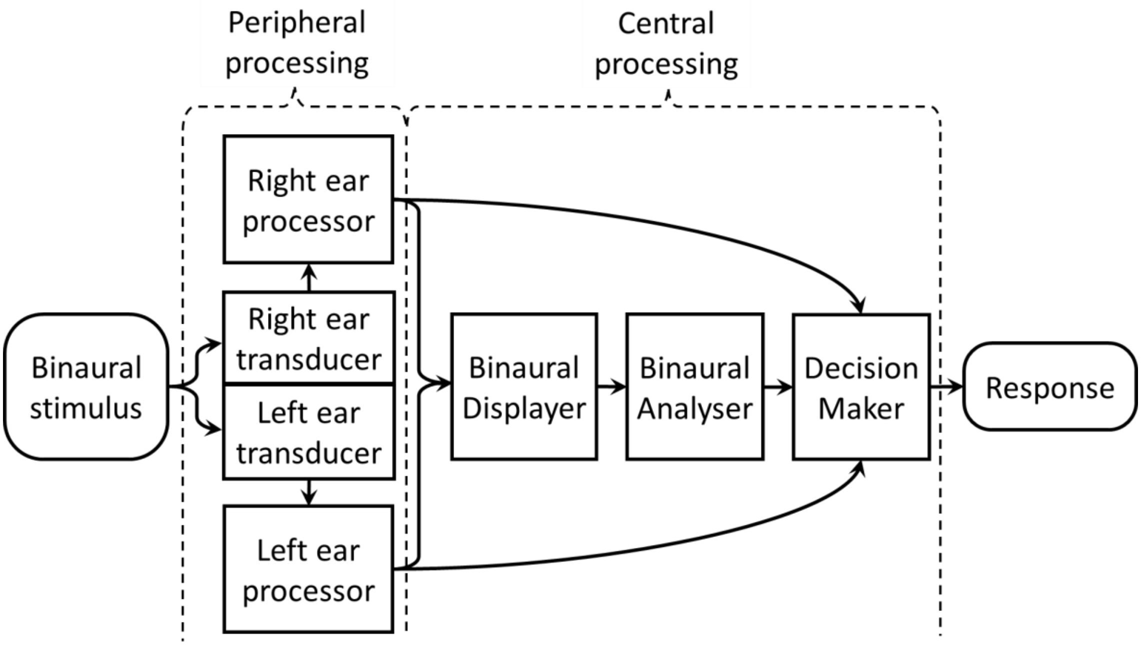


Figure 2.2: Colburn-Jeffress model schematic

The second stage is initially considered as an ideal central processor that optimally discriminates ITDs and ILDs. It operates an optimum combination of information from the two ears but does not necessarily make optimum use of all the information in the auditory nerve patterns. The ideal central processor requires that the time of events on a given fibre can be individually compared to the times of all events on the other auditory nerve. By restricting the set of allowed timing comparisons, a restricted central processor results in the model reflecting observations much better than an ideal central processor. The two constraints applied are that (1) each fibre is compared with only one fibre of equal CF, the two fibres forming a pair and (2) the information from each pair is limited to firings on the fibres that are almost simultaneous (within 100 µs) after a fixed internal interaural delay is applied. These constraints effectively restrict the central processor to an ensemble of coincidence-counting units, as originally proposed by Jeffress (1948). The latter constraint enabled Colburn to assume that the coincidence counting can be formalised as a Poisson process which leads to the statistical estimation of the predicted output of a coincidence counter. The resulting ensemble of estimations can then be used to predict discrimination and detection thresholds either by application of the Cramér-Rao inequality (the lower bound on the variance of an efficient estimator), or by direct prediction of performance that assumes that the decision variable is normally distributed.

2.2.3.2. Model validation for BMLDs

Colburn (1977) developed the model and applied it to the detection of tones in noise. He describes why the model copes well with various interaural parameters. Domnitz & Colburn (1976) had previously demonstrated how any model of detection based on interaural cross-correlation must correctly predict the dependence of thresholds on interaural target parameters for identical masker parameters in both ears. The predicted dependence on masker IPD is inherent to the structure of the binaural displayer. Since the coincidence-count in the central processor is basically a correlation measure, the effects of noise decorrelation achieved by combining independent noise sources and leading to an interaural correlation ρ will also be correctly predicted. The model works well as long as the brain can follow the fine structure of the sound. However, for frequencies above 1 kHz, firing patterns reflect more the envelope of the sound waveform and the model does not adequately reflect physiology. The predicted dependence on the spectral level of the noise is reflected in the model in that the predicted threshold for a single-fibre-pair is proportional (above fibre threshold) to the noise spectral level and in that detection threshold varies inversely with the number of fibre pairs contributing information, which itself increases with level, up to the point that all relevant fibre pairs are recruited. This explains the flattening following an initial growth of the BMLD with increasing masker spectral level. Finally, the predicted dependence on interaural masker amplitude ratio is a consequence of processing by single-fibre-pairs and of the number of pairs in use.

At a glance, a comparison of Figures 2.1 and 2.2 illustrates how similar in their general structure the Durlach (EC) and Colburn-Jeffress models are. They both represent the cochlea as a transducer that is frequency-selective across multiple channels. They both include three processing paths, two monaural and one binaural, which present decision variables to a decision maker. They both contain a binaural, central processor that takes as a key input the time difference (or coincidence) of arrival of a signal between the ears. As a result they both fail to predict data at higher frequencies where the peripheral auditory system does not encode the fine structure of sounds. Colburn & Durlach (1978) noted that Colburn’s auditory-nerve-based model can be regarded as a generalisation of the EC theory. However, the concepts underlying central processing in the two models are different, one working from the assumption of EC, the other from the assumption of CC. As a result, two branches of model development stemmed from the original models.

2.2.4. Extended models of BU based on the EC theory

More recent extensions of the EC theory (Culling & Summerfield 1995; Breebaart etal. 2001a; 2001b; 2001c) also used a restricted set of transformations that include only delay and attenuation.

2.2.4.1. Speech recovery in noise

Culling & Summerfield (1995) developed an extension of the EC model to accommodate vowel-identification data and explore the role of within- and cross-channel processes recruited in the separation of competing complex broadband sounds as a function of their interaural phase spectra. By employing competing vowels presented via two discreet bands of noise reflecting the first and second formants (F1 and F2), correct identification was easily achieved in dichotic presentation but the vowels were not identifiable in binaural presentations with differing ITDs. Thus, little evidence of cross-channel grouping with ITDs was found. Only when one of the vowels had its F1 and F2 bands decorrelated could that vowel be correctly identified. This suggested that the binaural system made a decorrelated signal more salient in situations where ITDs as a grouping cue for that same signal was ineffective. Frequency selectivity of one ear was modelled by a bank of gammatone filter whilst inner hair-cell transduction was represented by the Meddis (1986; 1988) hair-cell model. The signal was then processed through an EC stage that used a heuristic rule for selecting the optimum delay to apply (the minimum of the difference function within each band) and hence applied a different delay across bands. The above results supported Culling and Summerfield’s idea that, whilst localisation requires concurrent processing of ITD and ILD cues across frequency channels, the recovery of speech from noise by binaural unmasking exploits the interaural decorrelation introduced by differences in ITDs *only*.

2.2.4.2. Resolving the BMLD prediction issues at high frequencies

Breebaart et al. (2001a; 2001b; 2001c) produced a series of publications on an extension of the EC model motivated by new physiological evidence. In the medial superior olive, so-called excitation-excitation (EE)-type cells exhibit a discharge rate that depends on ITDs in that their response to a binaural stimulus is higher than to a monaural stimulus and is maximal when the stimulus matches the cell’s *characteristic delay*. The discharge rates resulting from EE interactions are usually modelled by a CC function. In the lateral superior olive and inferior colliculus, subgroups of so-called excitation-inhibition (EI)-type and inhibition-excitation (IE)-type cells are excited by signals from one ear and inhibited by signals from the other ear, which makes EI and IE cells sensitive to interaural intensity differences (IIDs), with a characteristic minimum IID. EI-type cells also exhibit ITD sensitivity which, together with their IID sensitivity, would support an EC theory account of BU. Given the above-suggested physiological basis for ITD and IID processing and knowing that the EC model had proved able to cater for a wider range of stimuli than correlation-based models (e.g. non-Gaussian noise, stimulus level variability or effect of signal or masker duration), Breebaart et al. (2001a) based their model on the EC theory. The model consists of two *peripheral processors* (outer and middle ear transfer function, filtering by the basilar membrane, transduction by inner hair cells and adaptation loops); a *binaural processor* as a two-dimensional array of EC operators (delays, attenuators and EI-type elements) taking input from each peripheral processor and followed by addition of internal noise on both monaural and binaural paths; and a *central processor* (the *decider* in Figure 2.1) that receives monaural and binaural decision variable from the binaural processor and produces a response. Although the model is rather complex, Breebaart et al. (2001b) justified its complexity by the fact that their model accounts for many binaural detection phenomena and that the wide coverage of empirical data would not be possible if any of the elements of the model was removed. The unification of ITDs and IIDs enabled adequate reflection of 'classic' BMLD data (N_0S_0 vs. N_0S_π or $N_\pi S_0$) dependence on f_0 above 1.4 kHz, which we had seen was problematic in both the original EC and Colburn models. The level and bandwidth dependence of classic BMLDs were also better predicted than by previous EC model variants.

2.2.4.3. Effect of temporal parameters

Breebaart et al. (2001c) examined model prediction dependence on temporal parameters. Correct predictions were found for $N_\rho S_\pi$ and $N_\rho S_M$ thresholds against the interaural correlation ρ of wide-band noise and for $N_\rho S_\pi$ versus ρ with a narrow noise bandwidth. Whilst N_0S_π thresholds were well predicted as a function of signal duration, predictions were poorer for masker duration. The model however correctly predicted trends in discrimination of time-varying ITDs and IIDs. Even the forward masking seen when the signal was presented a short time after masker presentation was reasonably well predicted.

2.2.5. Extended models of BU based on a correlation approach

More recent versions of models based on CC assume, as does the Colburn model, that signals are detected when the noise coherence (or interaural noise cross-correlation maximum) is reduced by the presence of the signal (Durlach et al. 1986; Lindemann 1986a, 1986b; Jain et al. 1991; Culling et al. 2001, 2006). Listeners are very sensitive to small reductions in interaural correlation ρ from unity, but less so when the reference correlation is lower.

2.2.5.1. Interaural correlation & critical bands, discrimination vs. detection

Durlach et al. (1986) employed a simplified correlation-based model of BU that applies a simple bank of critical band filters (that reject masking components distant from a narrow-band target signal) and adds an interaural delay prior to cross-correlation processing within each band. The authors explored how well such a model could reflect the results of correlation discrimination experiments. In experiments such as those reported by Gabriel & Colburn (1981), where interaural cross-correlation discrimination was measured as a function of noise bandwidth at various reference correlations (typically 0 and 1) , the EC model predicted data reasonably well at $\rho = 1$. However, the EC model failed to reflect the data for uncorrelated noise in that the assumption that more information acquired via processing a wider noise bandwidth would lead to better performance went against empirical evidence. Durlach et al. (1986) examined cross-correlation, ITD and IID just-noticeable differences (jnds) as a function of a range of parameters. Much of the experimental data was found to be consistent with a CC account of BU. Correct jnd predictions were obtained as a function of target IPDs or IIDs in the $N_0S_{\varphi s}$ configuration, as a function of level and IID of the noise or when varying signal and noise interaural correlation. However, when the authors converted the psychometric functions derived by Gabriel & Colburn for cross-correlation discrimination (percent correct versus ρ increments) into predicted psychometric functions for cross-correlation detection (percent correct versus ratio of signal energy to noise power per cycle), the predicted function slopes were significantly lower (3%/dB) than empirical data exhibited (4-8%/dB) for N_0S_π and N_0S_M. The authors remarked that the critical-band filtering that occurs for signal detection in noise does not appear to occur for correlation discrimination and that a close relationship between BU and correlation discrimination does not hold for the detection of a diotic signal in uncorrelated noise or in a coherent masker with only ILDs.

Jain et al. (1991) used a simplified version of the model of Durlach et al. (1986) to study analogous cross-correlation discrimination and signal detection experiments for diotic or uncorrelated noise. The model assumes critical-bandwidth filtering immediately followed by cross-correlation, i.e. without insertion of interaural delays given the restricted noise conditions. Whilst masking experiments indicate no interaction between critical bands, the degradation of cross-correlation discrimination at noise bandwidths exceeding critical bandwidths (so-called *supercritical bandwidths*) suggests interactions exist. To investigate these inconsistent findings, a spectral fringe was added to a narrow-band stimulus in a cross-correlation discrimination experiment. The particularity of such a *fringed-correlation discrimination* task is that the correlation change is limited to a narrow *target band* outside of which cross-correlation is held fixed with respect to the reference condition (for $\rho = 0$ or $\rho = 1$). This makes it analogous to a narrow-band target detection task at $\rho = 1$ and helps bridge the gap between the two types of experiments. Cross-correlation discrimination at $\rho = 1$ and N_0S_π detection appear to share a common mechanism. N_uS_0 signal detection performance does not relate to cross-correlation discrimination from $\rho = 0$ because binaural detection cannot be based on ITDs with an uncorrelated masker. When differences are restricted to a critical band through fringed-correlation discrimination, independent critical band processing paths consistent with detection can be used as a basis for the model. The authors' findings could also be explained by the two BU mechanisms proposed by Culling (2011). The first mechanism is sensitive to noise ILD modulations caused by addition of a target signal and exhibits susceptibility to frequency interference at N_0S_π. The second mechanism is mostly sensitive to ITD modulations and is relatively unaffected by across-frequency interference.

2.2.5.2. Binaural "sluggishness" and its relation to speech intelligibility

The binaural system is known to be somewhat "sluggish" in its response to stimuli with time-varying interaural differences, as though the information is temporally smeared within the auditory system (e.g. Grantham & Wightman 1978; 1979). Listeners are unable to track the instantaneous values of ITD or interaural correlation if they are varied with a frequency of more than a few Hz (Culling & Summerfield 1998). In model extensions that reflect binaural sluggishness, the instantaneous outputs of the coincidence-counting units undergo temporal integration using a temporal weighting function, which inherently causes temporal "sluggishness" because the duration of the integration window limits the resolution with which one can observe time-varying interaural differences of complex stimuli. Gabriel (1983), for instance, developed a black-box model in an attempt to unify ITD and IID discrimination phenomena with binaural detection phenomena for NH and

HI listeners. Gabriel's model incorporated separate time constants for processing ITDs and IIDs to reflect binaural sluggishness.

Although binaural sluggishness has been modelled with some success, a question is whether it is relevant to speech intelligibility. Using a detection task, Culling & Summerfield (1998) measured NH binaural temporal windows with equivalent rectangular durations ranging 55-188 ms. In considering the BU relation to speech intelligibility, Culling & Colburn (2000) reasoned that, if BU is indeed underpinned by cross-correlation, binaural speech intelligibility improvements over monaural conditions must be mediated by spectro-temporal variations in cross-correlation that mirror such variations in the speech. In a first experiment, the authors demonstrated that the discrimination of spectro-temporal patterns in noise (ascending or descending pure-tone arpeggios) was susceptible to binaural sluggishness. Indeed the binaural advantage of N_0S_π versus N_0S_0 conditions was found to drop dramatically with the rate of frequency change. In a second experiment, SRTs in noise were measured as a function of articulation rate in the same two binaural conditions. As articulation rate was doubled from a normal rate, the BILD was nearly halved (from 5.2 to 2.8 dB) and both conditions exhibited a 6-8 dB threshold increase. Culling & Colburn concluded that speech modulation frequencies useful to speech intelligibility (< 5 Hz) are not affected by binaural sluggishness.

2.2.6. Evidence supporting the EC-theory account of BU

Correlation-based models of BU account for binaural detection and discrimination in terms of the change in normalised correlation of the signals arriving at both ears. Van de Par et al. (2001) devised a detection experiment aimed at elucidating whether the correlation account of BU held in terms of the nature and precision of the normalisation required to reflect the data. Van de Par et al.'s basic argument was that the variability inherent to narrow-band stimuli typically used in detection experiments calls for a normalisation precision that may be unachieveable. Taking as an exemple the detection of a tone target in a narrow-band stimulus in the N_0S_π condition, the short-term power of the masking noise varies greatly in time. The first question is whether the detection of the small changes in correlation required to explain earlier binaural detection experiments could be achieved in the context of the correlation variability inherent to the baseline correlation, should normalisation be omitted. The authors demonstrate that, without normalisation, the typical jnds measured in correlation discrimination would be so small, compared to the standard deviations of the masking noise energy and of the resulting changes in baseline correlation, that detection would be impossible (the computed d' for the example was far too small). Normalisation is therefore required to account for binaural detection thresholds. The authors constructed an experiment aimed at maximising the normalisation precision required to account for the data. They compared the jnd in cross-correlation coefficient ρ between conventional N_0S_π stimuli and the same stimuli roved over a 30 dB range or presentation levels, which is a much wider range than that measured in conventional stimuli. The jnds measured for roved stimuli were found to be only slightly poorer than those of non-roved stimuli for a range of noise bandwidths and durations). Although this finding could be interpreted as evidence of normalisation operated within less than 20ms, even with normalisation taking place, the variability in stimulus power would have had to be reduced by more than 4 orders of magnitude to account for the measured jnds. This led the authors to believe that the binaural system probably does not normalise the stimuli presented to each ear, at least not directly. A proposed, much better suited, strategy would be for the central processor to subtract cross-correlation values of positive interaural delays from those of corresponding negative interaural delays. This would provide a measure of asymmetry in the cross-correlation function, which would effectively remove all external level fluctuations. The authors remarked that such processing could be achieved with EE-type elements (see Section 2.2.4.2). Insofar as the EC cancellation is supported by physiological models of nervous cells that receive both excitatory and inhibitory inputs, the authors concluded that their empirical evidence, together with earlier binaural detection data, favours the EC model over correlation-based models.

BLMDs from N_0S_0 are maximum at N_0S_π and typically measure 15 dB when a low frequency (250 Hz), narrow-band signal is presented in a broad-band noise. Culling (2007) compared for such stimuli the two competing accounts of BU. To do so, he conducted loudness discrimination experiments in a diotic and fixed broad-band noise, where the monaural power spectrum and the interaural correlation ρ of the narrow-band target were independently controlled. In the loudness discrimination task, both SNR and ρ were varied in the N_0S_π condition as well as in the 'Corr' condition that made use of a fringed narrow-band stimulus (as per Jain et al. 1991, see Section 2.2.5.1). In the N_0S_π condition, both reductions in ρ and increases in target energy contributed to a perceived increase in loudness with SNR. In the Corr condition, results at low reference SNRs (high reference ρ) were identical to those of the N_0S_π condition in that the same correlation change gave the same level of discrimination. At higher reference SNRs, however, the results of both conditions diverged, such that discrimination was always easier in the N_0S_π

condition, where the fact that additional energy was present in the target band (caused by adding in the signal) seemed to increasingly improve detection. The apparent increased effectiveness of the intensity-change cue thus suggested an interaction between monaural and binaural cues. In a second experiment, Culling set out to characterise such an interaction by observing how the monaural cues (represented by N_0S_0) combine with binaural cues (represented by Corr) to give rise to the N_0S_π outcomes. Discrimination was tested in pairs of ρ values. The d' values for N_0S_0 were so small that neither a direct sum (assuming the two cues are related) nor a vector sum (assuming independent cues) of N_0S_0 and Corr d′ values could account for a superior N_0S_πd′. A super-additive interaction between monaural and binaural cues was not deemed a parsimonious conclusion at that point. The results were, however, potentially consistent with an EC account of BU in that the amount of (effective) anticorrelated sound added within the target band would be greater in the N_0S_π condition (and a larger cancellation residue would ensue) than in the Corr condition. The difference in cancellation residue between conditions would increase with increasing (effective) SNR. A third experiment sought to null out either the interaural correlation cue or the cancellation residue cue, which underlies EC theory, by keeping one or other of them constant while changing the spectrum level in the target band between intervals. Fixing a cue makes it impossible to solve the task using it. In the signal interval, the spectrum level of the target band was elevated above the flanking noise by a spectral prominence between 0.5 and 2.5 dB. The fixed ρ condition was designed to foil the correlation cue, but it gave consistently higher scores than a diotic control condition and the improvement increased with spectral prominence (as predicted by the EC theory). In contrast, the results of the fixed-cancellation-residue condition, in which the anticorrelated signal energy was kept constant, were indistinguishable from those of the diotic condition. Culling concluded that the combined results of the second and third experiments favour the EC theory in that they offer a more parsimonious account of binaural signal discrimination at moderate SNRs. The author also remarked that such a conclusion was particularly relevant for discrimination tasks, such as speech perception, where the task is performed above detection threshold.

2.3. EC-based binaural models of SRM

The models presented in this section make use of the EC theory and are kept simple by omitting the peripheral pre-processing (modelled outer/middle ear, basilar membrane, and hair cells) and either work directly on the signals to predict SRTs and SRM (Beutelmann & Brand 2006; Beutelmann et al. 2009; Beutelmann et al. 2010; Wan et al.

2010; Wan et al. 2014) or from binaural impulse responses to predict SRM (Lavandier & Culling 2010; Jelfs 2011; Jelfs et al. 2011).

2.3.1. Beutelmann, Brand & Kollmeier (2006-2010)

Beutelmann & Brand (2006) set out to construct a functional model for the prediction of SRTs and SRM from the combination of the EC and SII models. The EC/SII model was validated against SRT data acquired in various speech-facing, single-SSN-interferer spatial configurations and in various room acoustics for NH and HI listeners. Whilst previous binaural models predicted release from masking, Beutelmann & Brand's model was designed to process signal wave forms (target + noise) directly. The model's filter bandwidth was fine-tuned to reflect the widening from monaural bandwidth to binaural equivalent bandwidth (Beutelmann et al. 2009). A revision of the model led to the binaural speech intelligibility model (BSIM) that provided an analytical expression of BU for arbitrary input signals and was more computationally efficient. The short-term BSIM (stBSIM) was an extension (Beutelmann et al. 2010) that enabled the model to deal with fluctuating interferers, as per the approach by Rhebergen & Versfeld (2005, see Section 2.1.2.3).

2.3.1.1. A binaural model for the prediction of SRTs with a single interferer

The EC/SII model (Beutelmann & Brand 2006) was inspired by an earlier model by vom Hövel (1984). Their model consists of two gammatone filter-banks (one per ear); an EC stage, a selection stage that retains the signal from either gammatone filter-bank (the better ear) or from the EC stage output, whichever has the highest SNR; a final gammatone resynthesis stage; and a final stage that applies SII-weighing to the signal to reflect frequency importance.

The gammatone filtering that was initially applied is based on the shape of the auditory filtering by the basilar membrane and was selected to minimise artefacts after resynthesis (Hohmann 2002). It makes use of thirty bands, each band width set to one equivalent rectangular bandwidth (ERB) and all bands spanning 146-8346 Hz, as per Glasberg & Moore (1990). Hearing loss is accounted for using threshold simulating noise (based on pure-tone audiogram data). Elevated thresholds are modelled by adding to the masker signal an internal Gaussian noise that is uncorrelated between the ears, as per Breebaart et al. (2001a, see Section 2.2.3.2). Along one of the two channels, The EC stage applies the attenuation (for equalisation) and the delay that provides the best SNR post-cancellation within each band. The variance of attenuation and delay errors is made to depend on actual attenuation and delay applied, similarly to vom Hövel (1984), who had demonstrated that improved BMLD prediction accuracy could be obtained that way. Thus, attenuation and delay are frequency-independent (as they were in Culling & Summerfield 1995; see Section 2.2.3.1), and so are their respective errors. The SII for 50% intelligibility is computed using the ⅓OB method and intelligibility scores are derived (Fletcher & Galt 1950) for the specific sentence material used, the 50% intelligibility point being calibrated from the anechoic, collocated-situation SRTs.

Assessment of the effect of artificially imperfect binaural processing on the final result was achieved via Monte Carlo simulations, a computationally expensive process. Computational efficiency was vastly improved in the BSIM revision of the model, where a simplified gammatone filter-bank (Hohmann 2002) is applied, addition of a pair of constant intensity values to the noise used in the ITD calculation replaces the Gaussian noise that previously simulated hearing thresholds and the iterative search method for optimal SNR is replaced with a two-step calculation of the optimum attenuation for a given delay. The revisions of the EC stage processing errors are mathematically equivalent to low-pass filtering of the cross-correlation.

A proof of concept for dealing with fluctuating maskers was provided in the stBSIM extension. This model extension operates a simple slicing of the signal into band-independent short time frames (of 12 ms effective length), followed by averaging of SRT predictions across time frames. Thus, the stBSIM approach does not take into account binaural sluggishness (see Section 2.2.4.4).

2.3.1.2. Model validation against NH & HI listeners' SRTs

For the validation of the EC/SII model against non-modulated noise conditions (Beutelmann & Brand 2006), SRTs were acquired for NH listeners via headphone simulations and following the Oldenburg sentence test (see e.g. Brand & Kollmeier 2002) with a bespoke measurement application. The head-related transfer functions (HRTFs) and acoustics were reflected by convolution of the signal with BRIRs for the three acoustic settings employed (anechoic, office, cafeteria).

Speech-facing SRM (SF-SRM) predictions from the EC/SII model were consistent with SF-SRM data, SF-SRM peaking at a 100° masker azimuth for NH listeners in anechoic conditions. The impact of reverberation was correctly predicted as a reduction of SRM. The effect was not dependent on the RT_{60} of the room, but more consistently reflected the useful-to-detrimental energy ratio D of the impulse response (IR). D_{50} is the ratio of energy (calculated in OBs) between the early part of the IR (first 50 ms) that is useful to speech intelligibility (and primarily contains the direct sound energy), and the later part of the IR that carries energy detrimental to speech intelligibility (e.g. Bradley &

Bistafa 2002). Inaccuracies in predicting the effect of reverberation was attributed to the treatment of early reflections being left to the EC process rather than explicitly separated as per vom Hövel (1984). The impact of HL was correctly predicted as a reduction of SRM, which was clearly asymmetrical about the co-located masker position when HL was asymmetrical. However, for a mild HL, the model tended to underestimate the speech-facing SRTs (SF-SRTs) by up to 3 dB and to overestimate SF-SRM by up to 3 dB in anechoic conditions. This was to be expected, since HL was only taken into account through hearing thresholds (see Section 1.2.2.3) and effects such as the reduced temporal resolution HI listeners exhibit (Elliott 1975) cannot be simply incorporated in a model such as the BSIM.

The correlation coefficients of observed versus predicted SF-SRTs pooled across room conditions were 0.98 once individual prediction errors (averaged across configurations) were removed. This operation confounded somewhat the effect of HL on collocated SRTs and that on SF-SRM. The authors did not provide separate correlation slope information per listener group, but it is expected that the slopes would have significantly departed from 1 for HI participants since the effect of mild HL on SRM was underestimated. Pooled across noise azimuths and with individual errors subtracted, the media correlation coefficient was also 0.98. Overall, and mindful of the model limitations, the model performance was very good.

To evaluate the BSIM model and stBSIM extension for different listener types, acoustics and modulated noise characteristics (Beutelmann et al. 2010), NH and HI predictions and SRT measurements were compared for stationary SSN, 20-talker babble and speech-modulated SSN conditions. Acoustic conditions were varied, not only by BRIRs being acquired in various environments (anechoic, listening room, classroom or church) but also by concurrently varying the listener distance from speech (3 or 6 m), from noise (2 or 4 m), or from a reflecting wall (far or very close) that could generate a strong noise reflection in the better ear, akin to a second, virtual sound source. Only three speech-facing spatial configurations were retained: collocated (close sources), with the noise separated at 105º (close sources) so as to maximise SRM and with the noise separated at 45º (sources further away) and reflected (in non-anechoic conditions) from 135º by the wall opposite from the noise. The latter condition was hoped to substantially worsen the SRTs but was found to result in intermediate SRM between the collocated (no SRM by definition) and 105º-separation conditions. Both NH and symmetrically HI listeners were tested. The HI listeners were split into two groups, a matched group with typical, mild HL gradually increasing with frequency and a group made of the remaining

HI participants that exhibited either very low or high HL, some with sharp changes in their audiograms. The effect of room acoustics dominated the results, the church's delayed and strong reflections most severely limiting SRM but also worsening SF-SRTs. The second, marked effect was that of SRM, strongest in anechoic conditions and weakened in HI listeners. A third strong effect of dip-listening in the speech-modulated SSN conditions was highest for NH listeners in anechoic conditions and was reduced by both reverberation (down to non-significant in the church setting) and HL. The dip-listening trend was flattened on average across all participants and even occasionally reversed for HI listeners by the combination of church acoustics and a virtual second noise source.

Comparing the data to BSIM and stBSIM predictions, SRM benefit was correctly reflected for the collocated and 105º separation but underestimated for the 45º conditions. The dip-listening benefit was also underestimated across the board and the most disrupting interferer (stationary SSN or 20-talker babble) was not always correctly predicted. The most disruptive church acoustics effect was not correctly predicted as being much worse than that of the other rooms. Although overall correlations of observed versus predicted SRTs established across all participants and spatial configurations as a function of acoustics and noise type yielded high coefficients (0.81-0.96), thresholds were generally predicted to be lower than they actually were and more markedly so as the effects of reverberation and noise modulation increased. The slope of the regression lines tended to exceed unity and more so as reverberation increased. Overall, The combined effects of room acoustics and spatial separation of sound sources was well predicted by the model, provided the influence of room acoustics on the noise dominated the results, because the interaural decorrelation of the noise by reverberation and virtual noise sources created by early reflections directly affected the noise cross-correlation function within the model. The fact that the model cannot account for reverberation effects on the speech signal may explain part of the overestimation of SRTs, which was particularly marked in the church conditions. 70% of the variance of the SRTs of hearing-impaired subjects could be explained by the model (and its use of audiometric thresholds). Another important source of variance may have been HI listeners' reduced temporal resolution, which the model does not simulate. Residual variance seen in NH listeners may be explained by attentional or cognitive sources of variability.

In steady noise, the BSIM predictions compared to the data resulted in a reduction of rms error for NH and HI listeners when compared to EC/SII predictions (both rms errors improving by 0.4 dB at 1.3 and 1.9 dB respectively). This may have been due to having changed the EC stage processing error in the BSIM revision, to be analogous to the low-pass filter found in physiological models of hair cells. This change indeed reduced the interaural fine structure correlation at high frequencies.

2.3.1.3. Fine-tuning by measurement of the binaural filter bandwidth

The effective binaural bandwidth (EBB) is measured as up to 4 times wider than the monaural bandwidth for a narrow-band target and depends on the measurement method (see e.g. Hall et al. 1983). For broadband spectra, the EBB is known to be dependent on interaural phase relation between target and noise within each band (e.g. Holube et al. 1998), but its measurement is also dependent on the selected filter shape (Kollmeier & Holube 1992). In order to establish which EBB would best serve the BSIM model and its gammatone filter bands, Beutelmann et al. (2009) made new experimental measurements for conditions with IPDs that had a strong frequency dependence. IPDs were shaped as a sine-wave of a given periodicity along a logarithmic frequency scale. Using that sine wave period (in octaves) as a parameter, SRTs were measured for the alternating $N_{+IPD}S_{-IPD}$ conditions and compared to the reference (homophasic, non-alternating) $N_{+IPD}S_{+IPD}$ condition where no BU was expected. Monaural SRTs were also acquired to assess of IPD distortion effects. As expected, the authors found little effect of periodicity on the reference or monaural SRTs. Data and predictions for alternating and non-alternating conditions were both found to converge beyond a 2-octave period. The rms error between model and data in the alternating condition (with varying period) reached a minimum of 0.5 dB (median across subjects) at a gammatone filter bandwidth of 2.3 ERBs. Provided that the filter bandwidth was set within the limits mentioned above, it appeared reasonable to Beutelmann et al. that binaural processing in each frequency band be considered virtually independent of the adjacent bands (i.e. the equalization parameters can be chosen independently). Thus, the model's assumption of cross-band independence was maintained by the authors. However, the authors did not provide an explanation as to why the 2.3 ERB filter bandwidth was not applied to the BSIM model revision.

2.3.2. Wan, Durlach & Colburn (2010-2014)

The model introduced by Wan et al. (2010) differs with that of Beutelmann & Brand (2006) in that it is much more faithful to the original EC model (Durlach 1963; Durlach 1972, see Section 2.2.1). It was therefore presented by the authors as an EC-model extension that could predict speech intelligibility in steady noise as opposed to BMLDs. It was also applied to multiple-masker situations and a recent, short-term version (STEC, Wan et al. 2014) was developed to deal with fluctuating noise and speech maskers, getting ever closer to predicting speech intelligibility in a cocktail party. Basing the extended model strictly on the EC model meant, however, that it was restricted to applications in anechoic environments.

2.3.2.1. Combining the EC model with SII weightings

A previous extension (Zurek 1993) of the EC model had demonstrated that by combining the EC model with SII weightings, the dependence of the intelligibility threshold on the angle of the masking noise (relative to the angle of the speech source) was well predicted for available SRT measurements. Directly applying the EC model, BMLDs were also found to be predictable from narrowband detection benefits, even with multiple maskers (Culling et al. 2004). In contrast with the Beutelmann & Brand (2006) model, the Wan et al. (2010) steady-state EC (SSEC) and (2014) short-time EC (STEC) models assume that every time sample of the filtered stimulus waveform is independently jittered and that jitters are applied independently in each frequency channel. Extensions to the (Durlach 1963; Durlach 1972) EC model present several alterations to the original assumptions. Firstly, time-varying jitters are applied to both interaural time delays and interaural amplitude ratios. Secondly, stimuli are processed by equalisation of the masker in each frequency band separately. Thirdly, the information is combined across bands using the SII. Fourthly, a full equalization of interaural level is allowed. The STEC model further assumes that the EC process in each frequency channel varies over time and that a sliding time window (20 ms-long and rectangular) overlaps 50% of the adjacent time windows, thereby preserving (according to Drullman et al. 1994) enough envelope information within each band to allow good speech intelligibility. The parameters that describe the Gaussian jitter statistics (zero means and [σ_ε,σ_δ] variance) are the same for all channels and are equal to the values chosen by Durlach (1972). The SNR retained by the decision stage within each frequency band is the highest of the SNRs delivered by the two monaural and the binaural paths. The SII value is calculated using a linear weighted combination of the SNRs between -15 and +15 dB from all the frequency bands, applying the frequency importance weights from ANSI (1997). The SII criterion parameter for the SSEC model was chosen for a specific type and number of maskers to match the reference condition (target and maskers collocated and in front) and used to predict all other spatial conditions for which the same type and number of maskers were spatially distributed. In contrast, The SII criterion parameter for the STEC model was selected to match the most spatially separated condition for a given type or number of maskers. Thus, the validation of the STEC model was limited to the modelling of SRM. Validation for temporal (dip-listening) and confusability (IM) factors was left for future publication.

2.3.2.2. Validation with multiple interferers

The extended model was able to predict speech intelligibility performance in a number of masking situations, whilst still remaining compatible with tone-in-noise detection conditions. The authors focused the model validation on anechoic data published by Hawley et al. (2004) and data with high reverberation contrast (low vs. simulated high reverb) reported by Marrone et al. (2008). Hawley et al. (2004) was chosen for their varying number (1 to 3), arrangement (symmetrical or not around frontal speech) and type (SSN, speech-modulated SSN, speech and reversed speech using the same as the target voice) of maskers and because the measurements had been acquired both binaurally and monaurally. Marrone et al. (2008) had covered low and high reverberation situations for binaural SRTs with symmetrically separated (by 0, ±15, ±45 and ±90º) two-speech-masker configurations (different voices), for binaural SRM from symmetrically separating two reversed-speech maskers (±90º) and for monaural SRM from symmetrically separating two reversed-speech maskers (±90º).

The SSEC best-fit SII criterion (Wan et al. 2010) for the Hawley et al. data varied in the 0.297-0.369 range for the various interferer types. The worst rms prediction errors (inflated by a factor 2.5 to 6) were found where IM was plausibly present (speech and reversed speech), and speech modulation of SSN doubled the rms prediction error found with SSN (0.7 dB). Model predictions were particularly good for a SSN masker within a narrow SII criterion range, but as speech-modulation was introduced, the SII criterion range had to be substantially increased, SII going as low as 0.239 for a single interferer and gradually recovering to SSN level as the number of interferers increased to three and the effect of dip-listening was reduced. With criterion adjustments, the predictions still matched the data very well. This could not be said for speech and reversed-speech maskers, particularly in the binaural, two or three-masker situations where complex spatial effects on dip-listening and release from IM caused the model to predict poorer SRMs (by as much as 6 dB) than those actually measured. In contrast, the monaural SRM predictions remained good (within < 2 dB), suggesting that the EC model treatment of ITD and ILD cues cannot be made to reflect spatial IM through adjustment of a single parameter. The SSEC model fit to the Hawley et al. data for a SSN masker was particularly good with 96% of the data variance accounted for (rms error about 1 dB), whilst it was poor for speech and reverse speech (rms errors 3.8 and 3.6 dB, respectively). Applying the STEC model (Wan et al. 2014) to the same Hawley et al. conditions appeared to provide improved predictions. Predictions indeed improved for speech-modulated SSN with two or three interferers to an accuracy comparable to that achieved for SSN. However, for speech and reversed speech, the improvement was only apparent because the authors used the most spatially separated condition as opposed to the collocated condition for the SII criterion adjustment. This change in prediction normalisation had little effect on the actual SRM prediction accuracy, it only shifted SRT predictions closer to the data for separated conditions, making collocated predictions poor (SRTs underestimated by up to 5 dB). After close inspection of the data, and contrary to the authors' conclusion, one could argue that STEC predictions were no better than SSEC prediction. In fact, they were not as good in some cases, such as the reversed-speech and speech situations when maskers are present in both hemifields, presumably making it harder for the listener to spatially release speech from IM. In these cases the STEC indeed has a tendency for underestimating thresholds (when compared to other separated conditions) and perhaps a lengthening of the sliding time window (to better reflect binaural sluggishness) would help better reflect spatial release from IM. The authors could have provided a perhaps fairer overall comparison of the two models had they adjusted the SII criterion to achieve best fit over all spatial conditions within a set rather than normalising predictions separately for each masker type. What seems clear is that using the SII criterion as a single handle is not the correct approach to emulate IM and its variation with spatial configuration, regardless of the model version.

For speech and reversed speech conditions, an F0 is present in the masker. For both the Hawley et al. (2004) and Marrone et al. (2008) data sets, such conditions were associated with an increase in the SRM, which appeared to be associated with a release of IM. Wan et al. (2010) compared the Marrone et al. data and SSEC model predictions in terms of SRM for these conditions. The SRM predictions from the SSEC model diverged from the data by as little as 2 dB with reversed speech, but as much as 11 dB with speech maskers. Application of the STEC model (Wan et al. 2014) to the same data only reduced the 11 dB SRM gap to 9 dB with speech maskers, improving predictions a little for collocated and small separation (15°) situations. The combined spatial effect on F0 discrimination and spatial release from IM may be the main cause of the doubling of SRM prediction errors. If that were the case, the conclusion that the SII criterion as a single handle fails to emulate these effects would be reinforced. The authors did not attempt to predict the Marrone et al. conditions with high reverberation. Presumably, such an attempt could have led to still significant prediction errors (although SRM will have reduced significantly), or, arguably, to fortuitously accurate predictions. The authors presumably felt that confounding effects would have made any interpretation of prediction accuracy debatable.

2.3.3. Culling, Lavandier & Jelfs (2010-2013)

Culling and colleagues were specifically interested in the degrading effect reverberation has on intelligibility of speech in noise or competing speech and how reverberation reduces SRM. They did not attempt to include in the model (nor did they attempt to predict) the more complex effects of reverberation such as reduction of dip-listening, reduction in spatial release from IM or degradation of F0-segragation. Instead, Lavandier & Culling (2010) initially set out to build a computationally efficient model compatible with architectural acoustic software. Their model of SRM makes use of BRIRs instead of taking signal waveforms as inputs, follows a strict (BE + BU) account of SRM and uses a simplified EC-theory-based formula (Culling et al. 2004; 2005) that expresses BMLDs as a function of the IPDs of target and interferer and of interaural coherence of the interferer. Since the EC theory assumes a process of cancellation of the interferer, the decorrelation of the interferer is considered the prime source of BU reduction by reverberation, whilst IPDs reflect well the effect of source separation on BU. Earlier models based on EC theory (Levitt & Rabiner 1967b; Zurek 1993) could not deal with reverberation as they did not take account of interaural correlation. The Lavandier & Culling model enabled accurate prediction of SRM acquired with a variety of simulated target and interferer azimuths and distances from the listener, room sizes and wall absorption. In a later revision, Jelfs et al. (2011) substantially improved the computational efficiency and resulting precision of the model by removing superfluous signal processing steps. The authors validated the revised model against a range of data sets from the literature, showing accurate prediction of both BE and BU. Lavandier et al.(2012) proceeded to further validate the model under more realistic conditions with multiple noise sources and real-room acoustics. They demonstrated that this allowed the generation of complex "intelligibility maps" from room designs. Culling et al.(2012) made use of the model to predict SRM for NH listeners and CI users with one speech-shaped interfering noise and validated the model further by demonstrating good predictions for both listener types in real rooms.

2.3.3.1. Model specifics

The Lavandier and Culling model separately takes account of BE and BU contributions to SRM and simply adds them up. It therefore has two paths (see Figure 2.3a). BRIRs for target and interferer are first convolved with a short (4.3 s) SSN sample, so that the resulting artificial binaural signal samples contain the effects of reverberation. The target and interferer samples are then passed through two ½ ERB gammatone filterbanks (one filterbank per ear, both covering 20-10 kHz), after which they are fed through each (BE & BU) path.

The BU path computes the BMLD (in dB) within each band from the following Culling et al.(2005) formula:

$$(6) \qquad BMLD\ (\varphi_s, \varphi_m, \rho) = 10log_{10}\left(\frac{k-cos(\varphi_s-\varphi_m)}{k-\rho}\right)$$

where, within each band, ρ denotes interaural interferer coherence (cross-correlation function maximum), (φ_s –φ_m) is the difference in target and interferer IPDs and *k* expresses the effect of EC-theory time and amplitude jitters (see equation 5) with variability (σ_ε, σ_δ) = (0.25, 105 μsec), as per Durlach (1972). Equation *(6)* effectively expresses equation *(5)* for situations where noise coherence is reduced by reverberation or by the presence of the target. It does so by replacing γ with ρ.

Cross-correlation is calculated within each band with a 100 ms exponentially tapering time window (|WAVE, Culling 1996). Coherence and IPDs are extracted by searching (within ± 5 ms) for the time delays for which the cross-correlation function reaches maximum. (φ_s –φ_m) is computed from the target and interferer delays multiplied by the angular frequency ω_0 of the centre of the band. The BMLD output is the result of averaging BMLDs calculated at four epochs (0.5, 1, 1.5 and 2 s) and it is zeroed if the returned value is negative (no binaural benefit). The broadband BU is established by integration of BMLDs over the frequency bands, after SII-weighting them (referred to as SII integration).

The BE SNR within each band is computed by calculating a cochlear excitation pattern (Moore & Glasberg 1983) for each ear (via 256 0.13 ERB bands covering 0-10 KHz). BE SNRs then go through SII integration to form a BE target-to-interferer ratio (TIR), the broadband SNR, as perceived by the better ear. BE (TIR) and BU are then simply added up to predict an *effective* (binaural) TIR for the BRIRs used, and hence, for a given set of target and interferer azimuths, sound levels, distances from the listener and for a given room size and colouration. The model concerns itself neither with the effect of material used, nor with the correspondence between SII and intelligibility. Indeed, the psychometric function linking the SII to intelligibility of a specific material corpus is assumed fixed. Thus, the model does not directly predict SRTs, nor does it need to. Providing SRTs can be extracted from the data, effective TIR predictions simply need to be offset, either to match a reference condition (e.g. a collocated situation to evaluate the SRM produced by separated conditions) or to match the mean of all the data from a set of conditions (e.g. varying room colouration and separation to evaluate the effect of reverberation on intelligibility and SRM).

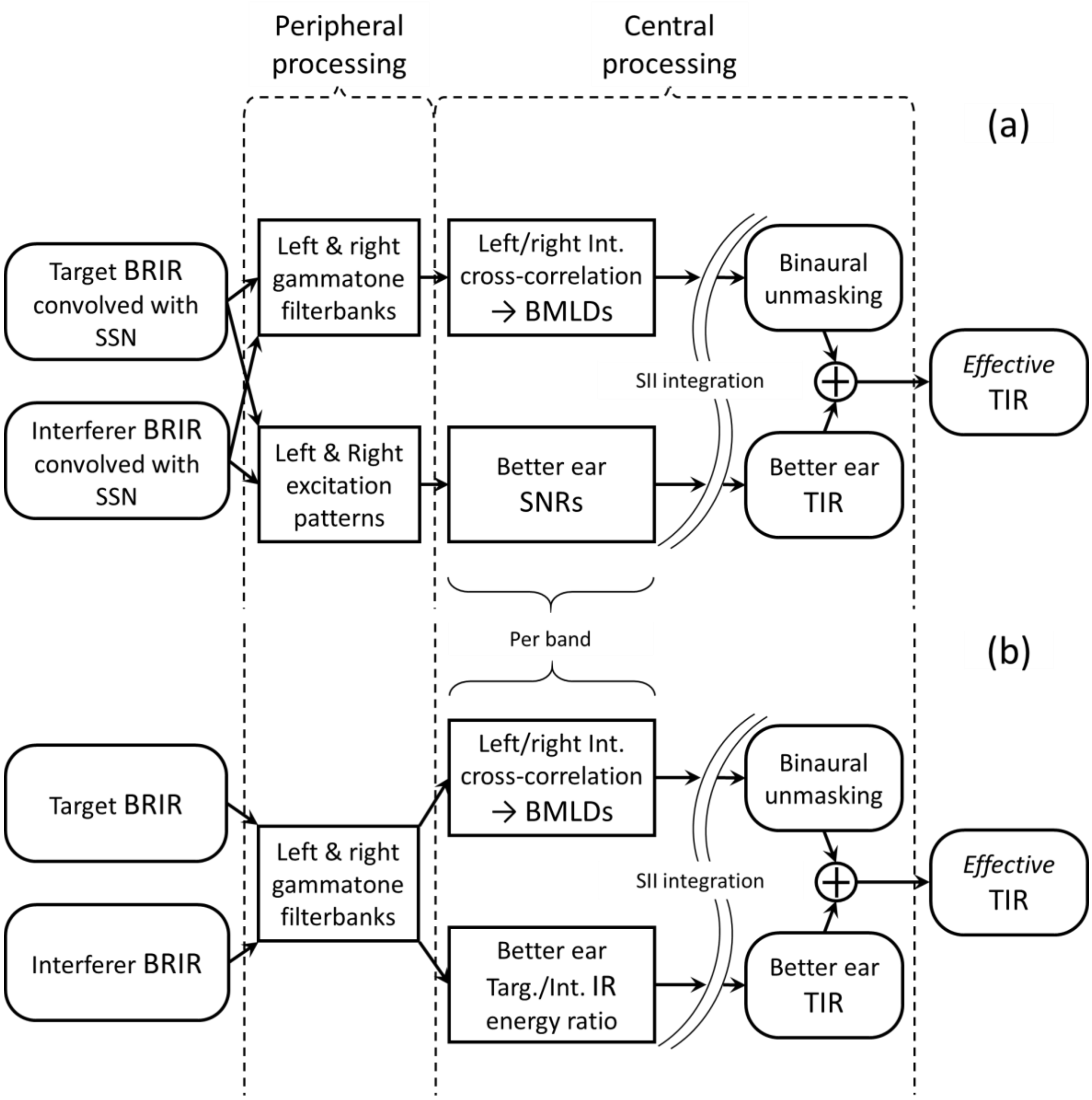

Peripheral processing
Central processing
(a)
Target BRIR convolved with SSN
Interferer BRIR convolved with SSN
Left & right gammatone filterbanks
Left & Right excitation patterns
Left/right Int. cross-correlation → BMLDs
Better ear SNRs
Binaural unmasking
Better ear TIR
SII integration
Effective TIR
Per band
(b)
Target BRIR
Interferer BRIR
Left & right gammatone filterbanks
Left/right Int. cross-correlation → BMLDs
Better ear Targ./Int. IR energy ratio
Binaural unmasking
Better ear TIR
SII integration
Effective TIR

Figure 2.3: Lavandier & Culling (a) and Jelfs et al. (b) model schematics

Jelfs et al.'s (2011) model revision recognises that convolution of signal samples with BRIRs was a superfluous step for steady-state maskers, as BU and BE-TIR could be calculated directly from filtered BRIRs (see Figure 2.3b). This was an essential simplification to make it possible for 2D speech-intelligibility prediction maps (in a given acoustic setting) to be generated within a reasonable timeframe for architects or acousticians. Cross-correlation is now operated directly on gammatone-filtered BRIRs to produce BMLDs, integrated in turn into a BU prediction. Furthermore, the BE TIR is computed directly from further processing of gammatone-filtered BRIRs as TIR can be accurately predicted from energy ratios between target and interferer BRIRs. The model returns 0 dB TIR for the collocated condition when it assumes target and interferer have equal power. Furthermore, the effect of multiple interferers can easily be modelled by simple concatenation of their BRIRs (joining them end-to-end). The effect of BRIR concatenation, other than adding their energy contribution to each band, is an averaging of the cross-correlation function between interferers.

2.3.3.2. Validation of the Lavandier & Culling model in NH listeners

Reverberation degrades speech-in-noise intelligibility not only by its direct effect on the target but also by affecting the interferer. Two experiments were designed by Lavandier & Culling (2010) to validate their model for this latter effect. The target was always kept anechoic to remove any effect of target temporal smearing. Because focus was placed on the BU degradation due to reverberation, the confounding effects of better-ear-listening were eliminated as far as possible by equalising the broadband stimulus levels between the ears. However, this meant that a residual BE within each band was still possible due to changes in room colouration coupled with source separation. Such changes could shift energy along the spectrum of an excitation pattern. The modelling of the rooms neglected the diffracting effect of the head, replacing it with two omnidirectional sound pick-up points 18 cm apart. The interferer was placed in simulated rooms of different sizes and colourations, and target and interferer were placed at different distances and azimuths from the listener. Experiment 1 had the interferer in a fixed position, deep in a room, whilst the target was at a shorter, fixed distance from the listener and had its azimuth varied. Both sources were in the frontal hemifield and such that SRM would cover its entire dynamic range, the interferer 16° away from the median plane that cut the room in half along its length. Predictions were offset so that data and predictions were equalised over all spatial separations in the anechoic interferer case. The correlation between predictions and data yielded a coefficient of 0.95. Over a 3.5 dB dynamic range of SRM, predictions were accurate within 0.5 dB, except in the highly reverberant condition where target and interferer azimuths almost coincided. There, the decorrelation of the interferer may have produced a little (0.6 dB) release from masking. Experiment 2 had a fixed azimuthal separation of 65° and the target was at a fixed distance from the listener. A wide range of conditions were tested for that involved room size, aspect ratio and wall absorption changes, as well as interferer azimuthal and distance shifts. The specific conditions chosen were intended to vary resulting phase difference and coherence so as to produce an adequate SRM range. The resulting SRM range spanned 4 dB, with a data-versus-prediction correlation coefficient of 0.97 and predictions all within less than 1 dB of the data. Computing SRM from BU predictions reduced the correlation, confirming that some room colourations led to a significant BE contribution to SRM, despite having equalised the interaural broadband rms. Overall, this was a compelling demonstration of how well the model could predict the effect of reverberant noise on SRM.

2.3.3.3. Validation of the Jelfs et al. model revision in NH listeners

Jelfs et al. (2011) validated the revised model against data from the literature. Validation was first demonstrated on the Lavandier & Culling (2010) data (experiment 2, SRTs with widely varied reverberation conditions for the interferer). The revised model marginally outperformed the original (R = 0.98). Jelfs et al. went on to predict data from various earlier publications (Bronkhorst & Plomp 1988 and Peissig & Kollmeier 1997, see Section 1.2.2; Hawley et al. 2004 and Culling et al. 2004, see Section 1.1.3). Both Bronkhorst & Plomp and Culling et al. had acquired anechoic SF-SRTs with ITD and ILD cues artificially separated, the former with a single interferer, the latter with up to three interferers. Correlation between data and predictions yielded coefficients of 0.86 and 0.95, respectively (for BE and BU effects separated, or combined where measured). The poorer correlation was attributed to a Bronkhorst and Plomp condition with the interferer was at 90º, where the *bright spot* effect can make the data very sensitive to the exact interferer azimuth (see Section 1.2.2.2). The Peissig & Kollmeier and Hawley et al. data sets covered SF-SRTs for various anechoic conditions with up to three interferers. Predictions for steady noise interferers correlated well with the data (R = 0.97 and R = 0.99 for the former and latter study, respectively).

Lavandier et al. (2012) validated the model for an anechoic target in a mixture of anechoic and reverberant, multiple interferers in large real rooms (130-240 m^2) and a lecture hall (500 m^2). In two of three experiments, cues made available in all the rooms were either ILDs-only or both ITDs and ILDs, with the target close by (0.65 m) and a single, near or distant interferer (0.65-10 m) and each source was either in front or at 25º azimuth. In the third experiment, simulating one of the meeting rooms with up to three interferers distributed in azimuth (within ±25º) and distance (up to 5 m), both cues were available with one, two or three interferers placed to the left of the target or two interferers symmetrically about a frontal target. The direct to reverberant ratio decreased when the noise source was moved away from the listener. The mixture of rooms and noise distances led to a wide range of reverberation. Experiments 1 and 2 were well predicted (R = 0.98, 12 conditions spanning 7 dB in SRTs) and so was experiment 3 (R = 0.95, 16 conditions spanning 4.4.dB).

2.3.3.4. Predicting the accessibility of a restaurant for NH and unilateral HL

Culling et al. (2013) discussed how the model naturally lends itself to architectural acoustics for room accessibility predictions. A variety of situations in a 9-table restaurant simulation were predicted for each sitting position, and for different table orientations, ceiling heights, wall absorptions, restaurant occupancy levels and listener head orientations with respect to an interlocutor sat at the same table. In addition to NH predictions, a mild unilateral HL was also considered. Amongst other findings, Culling et al.predicted the optimum listening strategy for NH listeners. Firstly, by chosing a corner table, the listener moved away from the bulk of noise sources. Sitting at one of those tables so as to face a wall (and the interlocutor) was predicted to provide them 6 dB more intelligibility than that found at the worst sitting position. Secondly, the listener could gain a further 3 dB HOB by orienting their head so as to both bring their better ear closer to their interlocutor and use the head shadow to shield that ear from interferers. Combined, the seating and head orientation strategies could boost speech intelligibility significantly. This would be a particularly welcome benefit for HI individuals and for HA or CI users. In the case of unilateral HL, half of the optimum seats for NH listeners would still favour the better hearing ear. The same would of course apply for unilateral CI users.

2.4. The Jelfs et al. model of SRM applied to CI users

2.4.1. Assumptions applied in Culling et al. (2012)

The Jelfs et al. (2011) model of SRM was first applied to the prediction of changes in speech intelligibility in noise for bilateral CI users and as a function of target and single-SSN interferer azimuthal separation in Culling et al. (2012). As seen in Section 1.2.2.6, two assumptions were considered reasonable for this modelling: firstly the assumption of equal effectiveness of the two CIs, secondly the assumption that CI users draw negligible benefit from BU in a speech-in-noise task. Thus, only the BE path of the model was processed.

Culling et al. acquired SRT data from CI users in an audiology suite with very high D_{50} (i.e. very good sound absorption by the walls). The model predicted very little SRM difference between that room and an anechoic environment. Coupled with a previous report indicating that the position of the microphone on the behind-the-ear (BTE) processor effects SRM little, but measurably (Aronoff et al. 2011), the authors thought appropriate to use head-related impulse responses (HRIRs), rather than BRIRs, from the front microphone of a Siemens Acuris hearing aid.

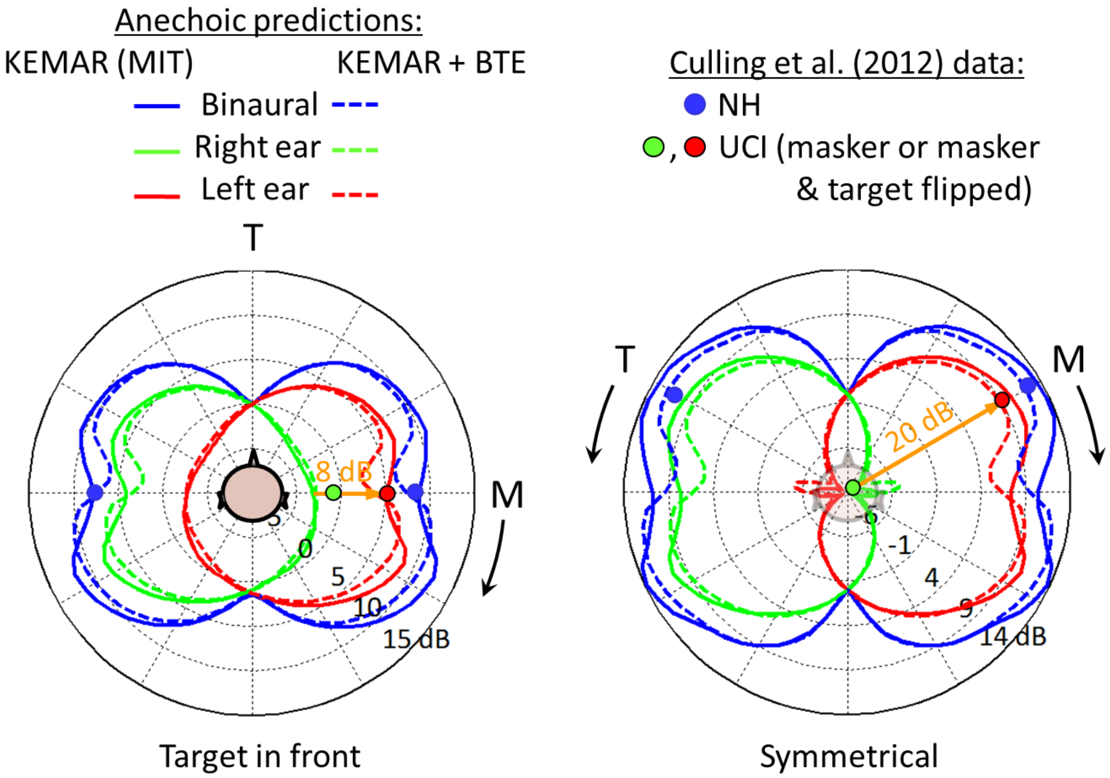
Microphone directionality was excluded from the study, since its effect on SRM is strong only when the interferer is placed in the rear hemifield (see Chapter 5). In any case, the authors wished to focus on the SRM for an omnidirectional microphone (or processor setting) and the spatial conditions employed in the study were restricted by keeping target and interferer in the front hemifield.

In the absence of representative SII weightings for CI users (and given the very wide variability such weightings would exhibit between users), NH weightings were also used. The authors argued that in a near-anechoic to anechoic environments, the better ear SNRs or TIRs (see Figure 2.3) vary slowly with frequency. There could be widely varying SNRs between adjacent bands only in high levels of reverberation (Lavandier & Culling 2010). Thus, it was adequate to predict CI users' SRM via 30 (NH) bands given that the predicted data was acquired in a sound-treated room.

2.4.2. Model validation for NH listeners and unilateral CI users

Given that the aim of the Culling et al. (2012) study was to demonstrate that the benefit of bilateral over unilateral implantation was, in selected circumstances, much larger than had been previously reported, the authors used the model to predict the results of previous studies (SF-SRM with interferer at ±90°), then test a situation that led to maximum predicted SRM difference between unilateral and bilateral CI users (symmetrical target and interferer separation at +/-60° and -/+60°).

Figure 2.4 illustrates the Culling et al. data in the context of model predictions that make use of the HRIRs acquired at MIT (Gardner & Martin 1995) with a Knowles Electronic Manikin for Acoustic Research (KEMAR) in anechoic conditions (solid lines), or HRIRs acquired from the front microphone of a Siemens Acuris BTE (dashed lines) fitted on a KEMAR manikin. As one can appreciate, there is very little difference at the orientations of interest between the two sets of predictions. The BTE predictions, compared to the MIT predictions, are rotated slightly forward by 5-10° (due to the change in microphone position with respect to the pinnae) and appear a little reduced as a result of that rotation (because the reference for SRM is the collocated situation). What is of most interest here, is that the model correctly predicts that for a symmetrical ±60° target and masker separation, the prediction difference between a left and a right-ear CI is 20 dB (all measurements were acquired with the help of unilateral CI users and the spatial configurations mirror-imaged). In other words, because the BE output of the model for a bilateral CI user is the outermost of the red and green lines, having two CIs rather than one can gain you as much as 20 dB SNR, the largest predicted and verified CI-user HS effect. Of course such large differences are only measurable in near-anechoic environments and with a single interferer.



Target in front

Symmetrical

Figure 2.4: Culling et al. (2012) data (circles) vs. predictions (lines) for NH listeners' and unilateral CI users' SRM

Thanks to their model of SRM, Culling et al. predicted that a combination of masker separation and head orientation away from facing the speech would lead to additive SF-SRM and HOB. One could derive from model predictions 2D maps of the head orientation leading to maximum overall SRM, when listening from any point in a virtual environment with fixed speech and noise sources. One of the key aims of this thesis is to validate such predictions for a range of spatial configuration.

2.4.3. Predicted effect of head orientation

The model was also used to predict how a speech-facing unilateral CI user positioned such that the noise was on their implanted side would experience a negative SF-SRM. In such a situation, as illustrated in Figure 2.5 (green lines), a unilateral CI user would have to turn their back to the speaker in order to achieve an optimal head orientation, which would be both socially unacceptable and preclude lip-reading. Many social situations are unfortunately not conducive to unilateral CI users changing their position in a room so as to bring the noise to their non-implanted side.

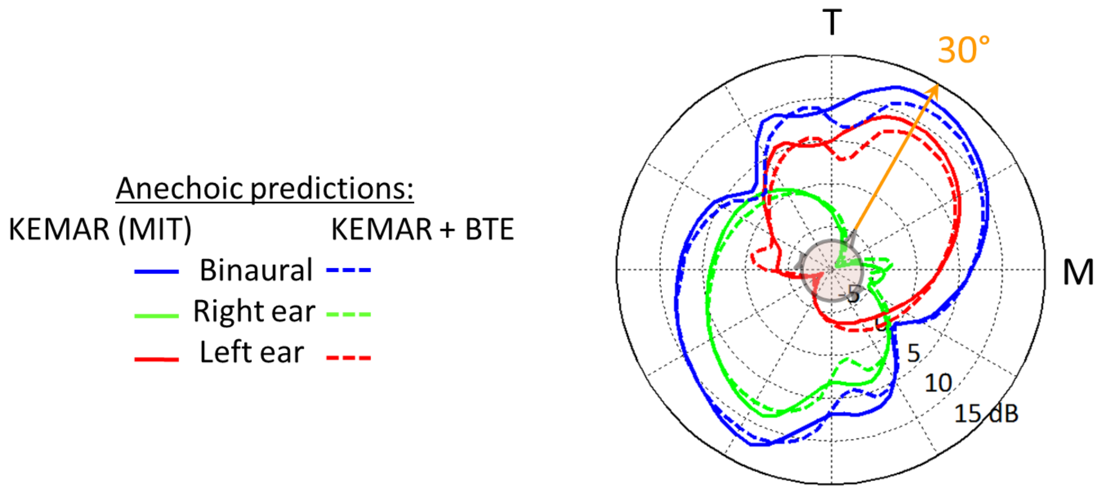


Figure 2.5: NH listeners' and unilateral CI users' (left or right ear) SRM predictions vs. head orientation from target in front and masker at 90º

In contrast, with equally effective implants, a bilateral user (for whom predictions are the outermost of the green and red lines) could easily use head orientation to shield one of their implants from the noise, thereby enabling BE listening and ensuing SRM. Predictions showed that a modest 30º head orientation (both socially acceptable and likely compatible with lip-reading) would enable a bilateral CI user to reap most of their available HOB, provided of course that the noise is sufficiently separated from the speech in the azimuthal plane.

As can be seen in Figure 2.5, no difference is found in the amplitude or orientation for the maximum HOB (4 dB) between SRM predictions drawn from anechoic HRIRs from MIT or the Siemens Acuris BTE. The model predicts that the largest HOB can be obtained when the masker is initially placed behind the (speech-facing) listener. This is illustrated in Figure 2.6. There, the maximum predicted HOB (at 65°) is 8-10 and 11-13 dB for CI users and NH listeners respectively. At the largest head turn that does not preclude lip-reading (30°, see Chapters 4 and 5), the difference in predictions between microphone conditions is small. The predicted 30° HOB is 5-6 and 8-9 dB for CI users and NH listeners respectively.

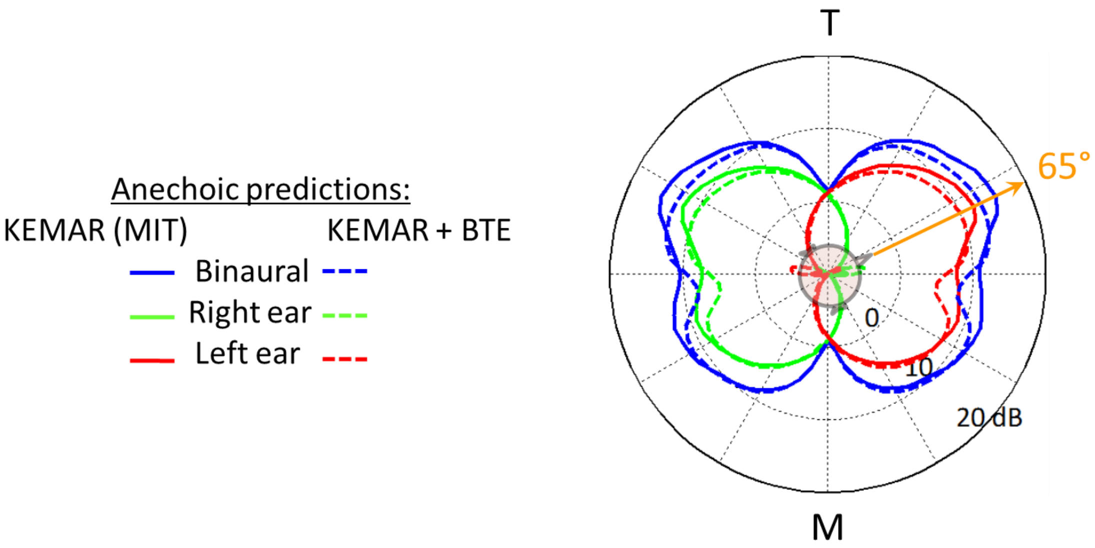


Figure 2.6: NH listeners' and unilateral CI users' (left or right ear) SRM predictions vs. head orientation from target in front and masker at 180º

2.4.4. Predicted effect of reverberation

The last section of Culling et al. (2012), discussed how the reverberation of different real environments may impact the HS effect for CI users. Simulations were made of a virtual-room equal in size and shape to the sound-treated room SRTs were measured in. The only changing parameters were the absorption coefficient of the walls and the floor. The predicted SRM increased nearly linearly with the absorption coefficient and with the corresponding RT_{60} reverberation times. However, it was further shown that RT_{60} itself is not a good predictor of SRM. The combined effects of wall, floor and ceiling absorption, room size, horizontal or vertical room aspect ratios and proximity of reflective surfaces in a variety of real environments were illustrated via acquisition of HRIRs in eleven rooms and SRM modelling for equidistant target and noise sources placed close to the listener (1m away) and on an azimuthal plane 1.3m from the floor. The rooms ranged from highly reverberant to near anechoic, from narrow (2 m) to wide (20m), from square to oblong (up to 5:1 aspect ratio) and with low (2.5 m) to high ceiling (6 m). In contrast with the output of the simulations above, varying multiple parameters highlighted how poor a measure of reverberation RT_{60} is when predicting SRM. Indeed, high levels of SRM could be obtained in both a sound-treated room and two cafeterias when their RT_{60} ranged 60-900 ms. Conversely, a university teaching room, a foyer and a wide corridor all yielded moderate levels of SRM whilst their RT_{60} spanned 200-1200 ms. Here, as previously seen in Lavandier's and others' work, the direct-to-reverberant ratio would do a much better job at predicting SRM than the RT_{60}. Culling et al. closed their discussion by highlighting how SRM will always be relatively high when the noise source is close by. As the noise source moves away from the listener to distances that are large compared to the circumference of the head, ILDs diminish (SPL following the inverse square of the distance, i.e. dropping 6 dB every distance doubling) and the BE benefit diminish with them. With reverberation, SRM diminishes further as moving the noise away reduces the direct-to-reverberant noise level ratio.

Given what we have learned in this section, the model predictions used in the following three experimental chapters will make the same assumptions as those justified by Culling et al. (2012) and will primarily make use of MIT HRIRs or BRIRs acquired in the test room(s).

3. Audio-only pilot experiments with young normally hearing adults

In the previous chapter, we saw how the Jelfs et al. model of SRM (Jelfs et al., 2011) generates predictions of SRM by summing its BE and BU contributions. Combined, the two contributions account for the two cues associated with SRM (Bronkhorst & Plomp 1988) and the two corresponding mechanisms involved in the separation of competing sounds (Culling et al., 2004; Hawley et al., 2004; Plomp, 1976). The BE contribution is derived by working out the SNR at the better ear in each spectral band whilst the BU contribution is derived from the EC theory (Culling, 2007; Durlach, 1963,1972) and the resulting BMLDs. The model was validated in previous studies (Culling et al., 2012; Jelfs et al., 2011; Lavandier & Culling, 2010; Lavandier et al., 2012) for binaural or monaural SRM predictions for a wide range of listening situations. Van Hoesel & Tyler showed that bilateral cochlear implant (BCI) users benefit from SRM (Van Hoesel & Tyler, 2003). Making use of the model, Culling et al. (2012) demonstrated that some spatial configurations led to much larger benefits of bilateral cochlear implantation than had been previously demonstrated. However, most experimental studies have only considered SRM in a fixed-head situation (e.g. Culling et al, 2012; Laszig & Aschendorff, 2004; Litovsky et al, 2006; Loizou et al., 2009; Lovett et al, 2010; Schleich et al, 2004; Van Hoesel & Tyler, 2003) and a majority of them considered solely speech-facing situations.

The model can inform studies of more natural, free-head situations by predicting SRM as the sum of (1) the effect of separating the masker from the target when the listener faces the target (SF-SRM) and (2) the effect of orienting the head away from the speech (HOB). A HOB is predicted to occur when target and masker are separated. This is due to the head-shadow effect. The model predicts that a HOB can be obtained by both unilateral and bilateral CI users. Since no work had been reported on normally hearing listeners' HOB and head-orientation strategies to date, a normal hearing baseline study was required. Our aims were threefold: firstly, to confirm that normally hearing listeners could reap the predicted benefits of head orientation when speech and masker are spatially separated; secondly, to design a free-head paradigm that would motivate spontaneous head orientations when listeners attend to speech in noise; thirdly, to establish how effectively normally hearing listeners make spontaneous use of HOB. In this chapter, we focus solely on responses to audio-only presentations of speech in noise with a single speech-shaped masker.

3.1. Experimental approach

Brimijoin et al. had previously combined SRT and spontaneous head orientation measurements (Brimijoin et al. 2012, see Section 1.4). Instead, we opted here for clearly separating the two. On one hand, SRT measurements would allow us to objectively confirm model predictions that head-orientation away from facing the speech is indeed beneficial when speech and masker are spatially separated. On the other, a free-head paradigm would separately enable observation of head orientations and strategies adopted by listeners. In order to give listeners ample opportunity to make use of head orientation in the free-head paradigm, we chose to present long clips with gradually diminishing SNR. At the start of a run, the SNR would be high, such that listeners could follow the content of the clip with ease. The expectation was that as SNR would approach the SRT when facing the speech, listeners would increasingly be motivated to make use of head orientation. This, we hoped, would constitute a stronger manipulation than presenting short sentences at or around the speech-facing SRT, as Brimijoin et al. did. We indeed suspected that presentation of short sentences was the reason why Brimijoin et al. had not succeeded in observing spontaneous head orientation away from the speech with normally hearing listeners (pers. comm).

We also made a point of ensuring that we would not compromise the undirected nature of the behavioural experiment. To that end, care was taken to ensure that no reference was made to head orientation until such observations were completed and that nothing in the lab could lead the listener to think orientation was one of our objects of interest. Consequently, each participant started their session with undirected, free-head listening tests. After these measurements were completed, instructions could be given about the potential benefits of orientation. A complete set of objective SRT measurements followed the first behavioural experiment. Finally, if time allowed, a second free-head experiment was run, this time in a directed manner.

3.2. Hypotheses and choice of spatial configurations

Results from the Jelfs et al. (2011) model of SRM were used to identify the spatial configuration that would yield maximum benefit of head orientation away from the speech. The model was fed with the MIT anechoic HRIRs (Gardner & Martin 1995). Jelfs (2011) used this method to create a 2D map of predicted maximum speech-in-noise intelligibility benefit achieved through head rotation away from the speech direction as a function of the listener's position in a 10 x 6 m virtual, anechoic room.

This map is reproduced in Figure 3.1 and was generated by subtracting the SRM for facing the target from the SRM achieved through optimal orientation. The maximum HOB was predicted when the listener was positioned exactly between the speech and the noise source, in other words with target and masker placed at 0° and 180° azimuths respectively (the T_0M_{180} configuration). Figure 3.2a shows the HOB in the T_0M_{180} configuration. The model predicts that for normally hearing listeners there is a benefit of up to 15 dB at ±65° head orientation in anechoic conditions.

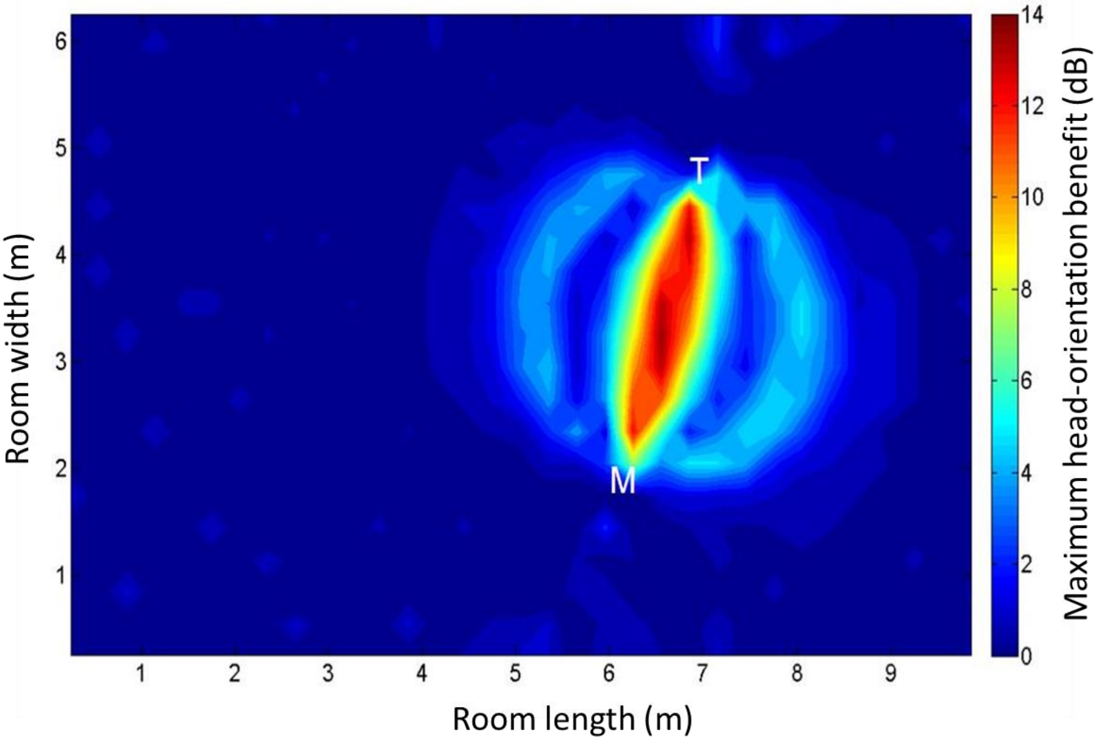


Figure 3.1: 2D map of modelled, maximum HOB in a 10 x 6m virtual room with one target (T) and one masker (M) – reproduced (Jelfs, 2011)

This pilot experiment examined whether normally hearing listeners can and do benefit from using appropriate head orientation. It thus had two objectives, first, to confirm the predictions of the model through fixed-head objective measures of SRT and second, to test whether listeners spontaneously make use of the optimum head orientations both predicted by the model and observed in the objective SRTs. 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 scan their head looking for improvements in SNR. Second, they may localise the sources that are present and then predict, from their analysis of the auditory scene, the optimal head orientation. Third, Brimijoin et al. (2012) suggested that 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. Indeed, head orientation alone affects target level at the ears, with a maximum level found for the 60° head turn that favours the ear attending to the target. Finally, perception of the masker in one hemifield may influence the listener to move their head away from it, which the model predicts would worsen their speech intelligibility.

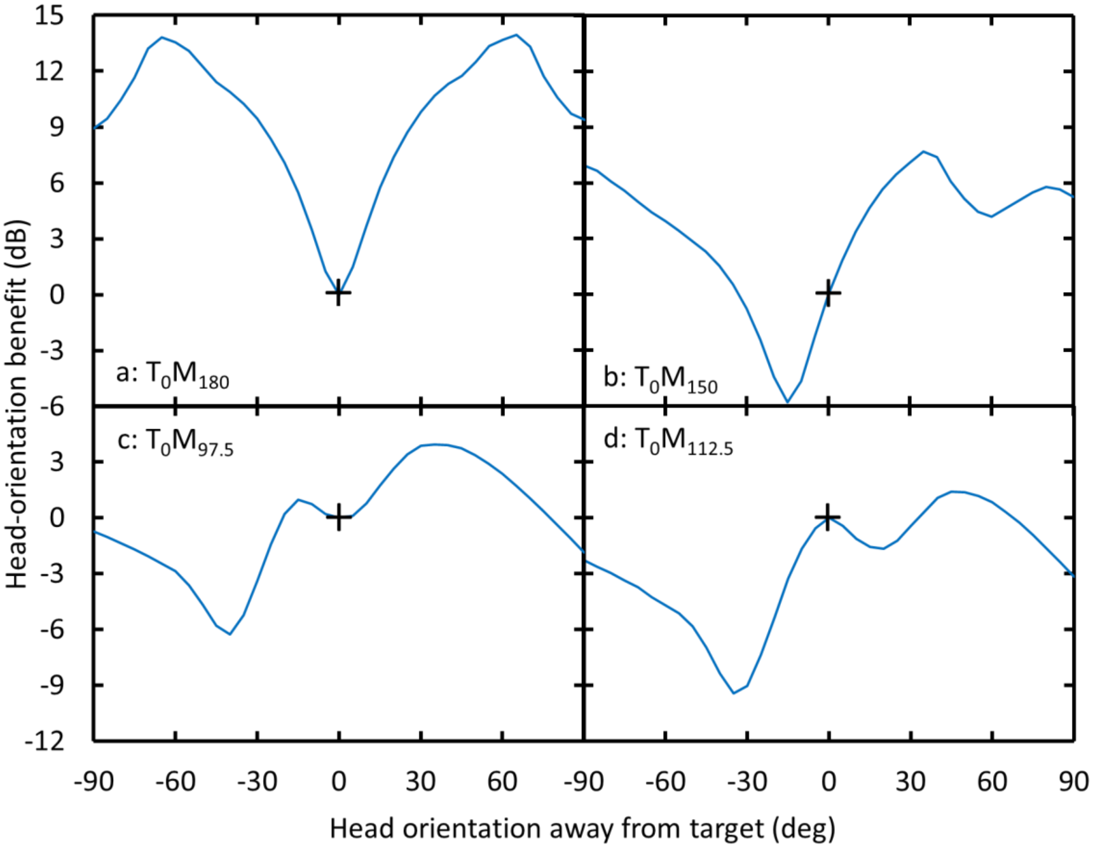


Figure 3.2: Predicted HOB in the four spatial configurations

In the T_0M_{180} configuration, the model predicts that listeners can obtain a HOB by orienting their head either way, but the head orientations predicted to provide maximum benefit (±65°) almost coincide with those of peaks in target level (±60°), so other conditions are needed to differentiate these possibilities. In the T_0M_{150} configuration, speech and noise were placed at 0° and 150° clockwise respectively. A large benefit was predicted for head rotation one way and a loss the other, as illustrated by the model predictions plotted in Figure 3.2b. If listeners can analyse the scene, one would expect listeners to rotate their heads to the right rather than to the left of the speech facing orientation. The orientations providing maximum benefit (30°) and maximum target level (60°) are further apart and it was hoped that this would help determine whether listeners aimed to maximise target level or SNR. The $T_0M_{97.5}$ configuration was chosen because the model predicted a benefit both ways, but with a larger benefit to the right than to the left (Figure 3.2c). Again, if listeners are able to analyse the situation and predict the best orientation rather than scan for it, one would expect them to turn only in the direction of maximum benefit. The $T_0M_{112.5}$ configuration was chosen as a control, where head rotation in either direction away from the speech direction would initially lead to a loss in intelligibility (Figure 3.2d). If listeners scan for SNR, they may be inclined to return to facing the speech after exploring surrounding orientations. However, if they can analyse this auditory scene then they might move to the HOB peak at 45°.

Listeners may be found to not move their head and to face the speech irrespective of where the masker is positioned. This could stem from social convention or it could be that they misinterpreted instructions given in a task or that they simply did not think of making use of head orientation.

3.3. Materials and methods

3.3.1. Participants

20 participants were recruited from the Cardiff University undergraduate population as well as the general population. They ranged from 19 to 50 years old, averaging 25 and all had normal hearing. The first 4 performed a preliminary experiment that allowed refinement of protocols and instructions. The data for the last 16 was retained and is presented herein.

3.3.2. Laboratory setup

A new audio-visual laboratory was developed to facilitate this research. A 3.2 m × 4.3 m sound-deadened room was equipped with a 3-m diameter 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 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. The RT60 of the sound-deadened room was measured 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 typically be at the centre of the speaker array. Control of the experiments could be achieved either from a computer station in the room or from an external control room. A Microsoft Lifecam 5000 video camera was fitted on the ceiling exactly above the listener's head so that either covert or overt video recording could be made of the listener's head orientation.

3.3.3. Free-head listening task materials & protocol

We did not to employ any head-attached metrology equipment for this experiment such as the infrared system used by Brimijoin et al. (2012), an accelerometer/gyroscope system or a magnetic field sensing system such as a Flock of Birds™ (Ascension Corp.). We felt these overt devices would potentially give our listeners too much insight into our purpose. Instead we opted for covert top-down video recording of the listener's head orientation. The resulting video recordings were post-processed in a semi-automated procedure making use of the Matlab mouse pointer function. Over two passes, an operator tracked the locations of the top of the listener's head and the end of 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 +/- 5° providing head orientation did not vary wildly, which was judged satisfactory for our purpose.

The material used for this experiment consisted of four 4-minute-long speech clips. They were speeches by President Obama obtained from the White House official site. All four clips exhibited consistency of speaker, speech flow, complexity and level. They talked of US internal or international affairs in a manner 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 FIR filter that was based on the calculated excitation pattern of the speech material (Moore & 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 0 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 his/her 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 during the task*". This ensured that the listener's head remained in the centre of the speaker array and that they 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 do so.

Having acquired the undirected behavioural data, we then informed the listener that head orientation might be beneficial and repeated the test after completion of the SRT runs. Listeners were 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 behavioural experiment. Whenever time allowed, a set of these directed runs was performed, re-using the same speech material.

Since the speech material varied in quality between clips, each clip was allocated to and always presented with its associated spatial configuration. This ensured consistency of speech within a configuration and possible subsequent statistical treatment of the data. Rotation of configuration against material was not judged essential here since our main interest was finding out whether people would naturally rotate their heads and what their final head position would be. However both spatial configuration and their associated speech clip were rotated across participants.

3.3.4. Fixed-head SRT task materials & protocol

For each condition the SRT measurement involved a list of 10 sentences from the IEEE corpus (Rothauser et al. 1969). These are semantically plausible but unpredictable sentences such as ‘a LARGE SIZE in STOCKINGS is HARD to SELL.’, that each contains five nominated key words (shown in upper case), one of which has two syllables. A 1-up/1-down adaptive threshold method (Plomp & Mimpen 1979b) previously used in Culling et al. (2012) was employed. Speech from speakers DA or CW was consistently presented through the 0° azimuth loudspeaker. The number of sentence lists required was such that both DA and CW lists needed to be used. The continuous speech-shaped noise spectrally matched the speech material and 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 level was kept constant at 70 dB. The staircase started at -20 dB SNR. The listener was required to repeat as many of the five key words as they could and the experimenter input the number of correct words. Since the starting SNR was low, the first sentence was presented again with speech level increased in 4-dB steps until the listener correctly repeated at least 3 of the key words. From then on the adaptive phase started, the sentence was changed every trial and speech level stepped up or down by 2 dB when the listener repeated correctly less or more than half of the keywords respectively. The last 8 SNRs computed were then averaged to calculate the 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.

8 azimuthal orientations of the head were used to construct a partial map of SRTs surrounding the speech-facing orientation. The exact azimuths chosen aimed at confirming the neighbouring maxima and minima found in the predictions (Figure 3.2) and so for each of the four target and masker spatial configurations selected. Hence a total of 32 IEEE sentence lists were used to cover all head orientations and spatial configurations. Speech material was kept in the same order for all participants. Trials were grouped in blocks of 8 to cover each of the selected head orientations within a given spatial configuration. The order of the blocks and hence the spatial configuration was rotated for each new participant. Within each of the four blocks, the head orientations were also rotated.

3.4. Results

3.4.1. Undirected task head movements

Of 64 undirected trials, the data from 62 were retained since one participant on two occasions made use of their hands to either block the masking noise or reflect the target speech into their ear. The most significant finding was that in 60% of the trials (37 of 62) listeners spontaneously moved their head away from the speech 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. Figure 3.3 shows an example time plot of undirected head orientations adopted by a few representative participants. The filled circles at the end of each track correspond to the clip time at which listeners flagged losing track of the speech. These points are therefore subjective measures of SRT achieved at the final head orientation. The head tracks are displayed in the context of model predictions (pink bands). The predictions were moved along the subjective SRT axis so as to equalise the means of subjective SRTs and predictions across all spatial configurations.

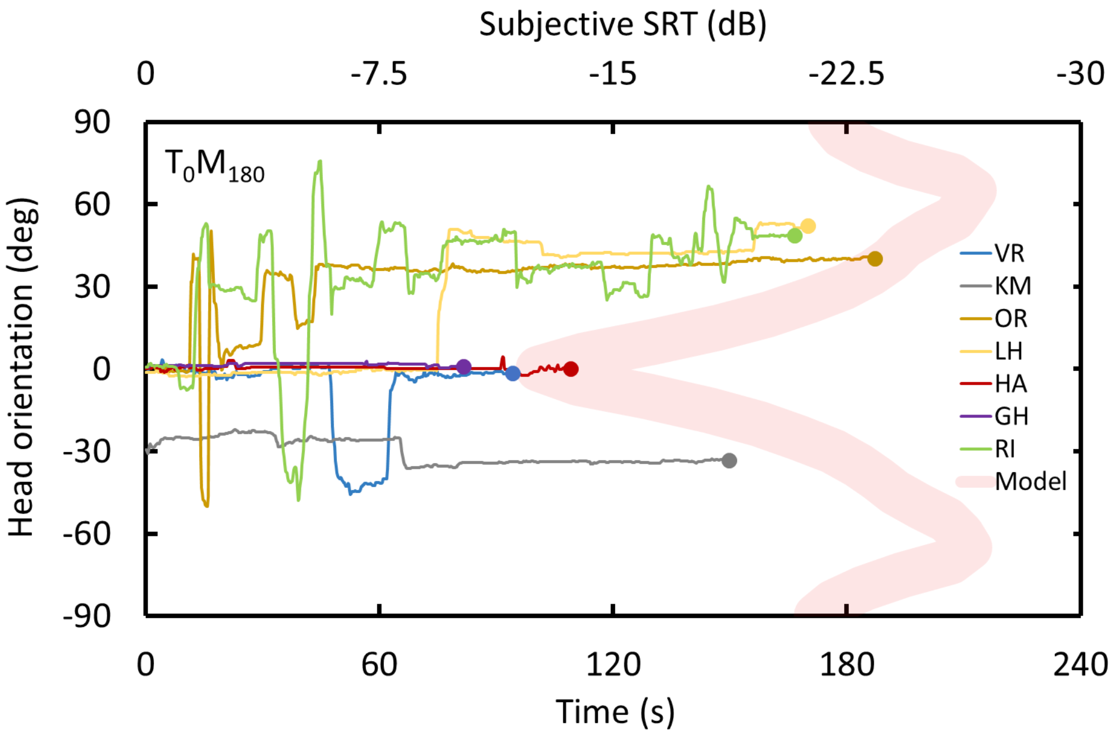


Figure 3.3: T_0M_{180} head-orientation example tracks (labels are participant codes) against absolute clip time (lines) and subjective SRTs (filled circles) achieved pre-instruction

The first noticeable point is that, despite individual differences, subjective SRTs reached through head orientation broadly follow the model predictions. 5 participants did not turn their heads, 6 turned their heads to the right and 5 the left. Of those who did move, some did so seemingly 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 seemingly more gradually, thereby gradually improving, but not necessarily reaching the optimum head orientation. Others jumped more or less straight to the most beneficial orientations centred on ±65°. It was not possible to categorize participants, however, since we found no evidence that they do not belong to a continuum.

Example head-orientation tracks in the three asymmetrical configurations (T_0M_{150}, $T_0M_{112.5}$ and $T_0M_{97.5}$) are showed in Figures 3.4-6. As in the T_0M_{180} configuration, some participants did not move, some moved seemingly gradually or erratically, sometimes going as far as achieving a loss after having passed through a benefit region and others performed best by seemingly jumping straight to the region of maximum benefit. Here too, subjective SRTs broadly followed the model. In 20 out of 48 trials, listeners remained facing the speech. Those who moved turned equally frequently to the left (14 trials) or to the right (14 trials).

The amount of head movement over each run was computed as the average unsigned, wrapped head orientation. An ANOVA operated across all four spatial configurations resulted in a marginally non-significant effect of configuration on head movements ($F(3,42) = 2.55$, $p = 0.069$). Head movements were therefore only marginally larger in the T_0M_{180} and $T_0M_{97.5}$ configurations. Head movements may have been more strongly motivated in the T_0M_{180} configuration by the larger HOB experienced (when listeners ventured away from facing the speech) and in the $T_0M_{97.5}$ configuration by the stronger masker level in the ear almost facing the masker when the head faced the speech. An ANOVA operated across configurations revealed a significant effect of configuration on the subjective SRTs ($F(3,42) = 10.06$, $p < 0.01$). This was to be expected as SF-SRM is known to vary widely with masker separation. Despite the HOB contribution to SRM being strongest in the T_0M_{180} configuration, the mean subjective SRT was on average 4.5 dB lower in the asymmetric configurations. Had listeners reached optimal HOBs, no effect of configuration on subjective SRTs should be found. This further illustrates that listeners were poor at spontaneously reaching optimal HOB. Note here that one listener did not have time to complete all runs and hence was excluded from the statistical analysis.

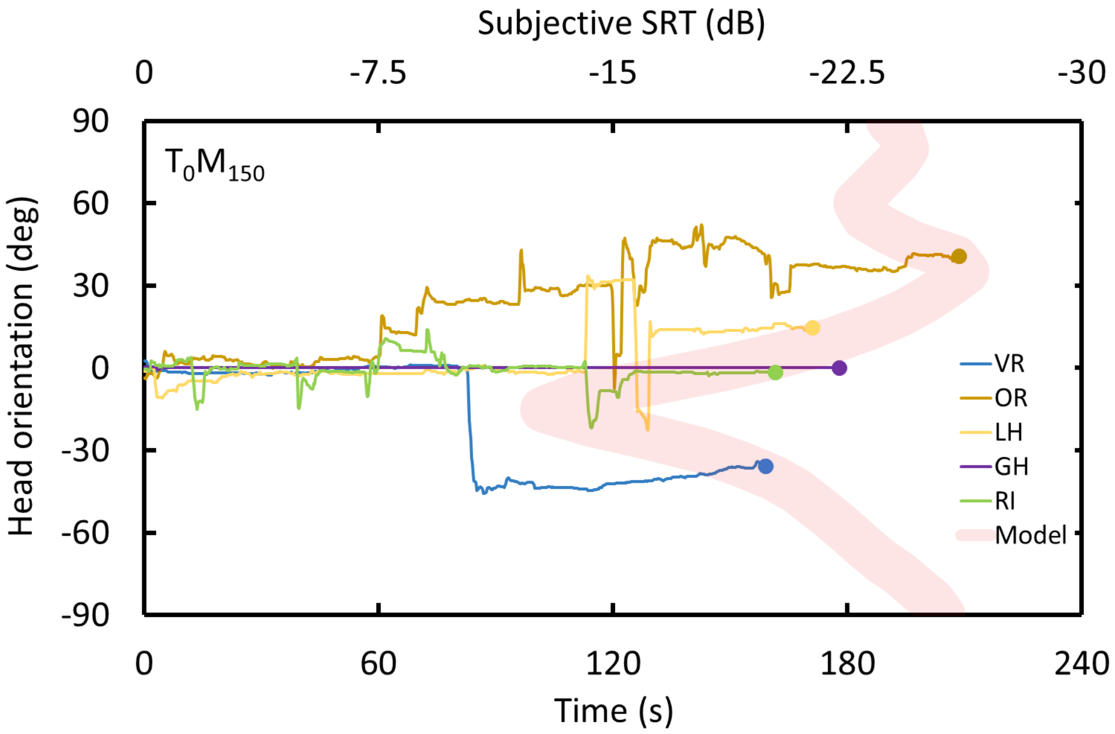


Figure 3.4: T_0M_{150} head-orientation example tracks (labels are participant codes) against absolute clip time (lines) and subjective SRTs (filled circles) pre-instruction

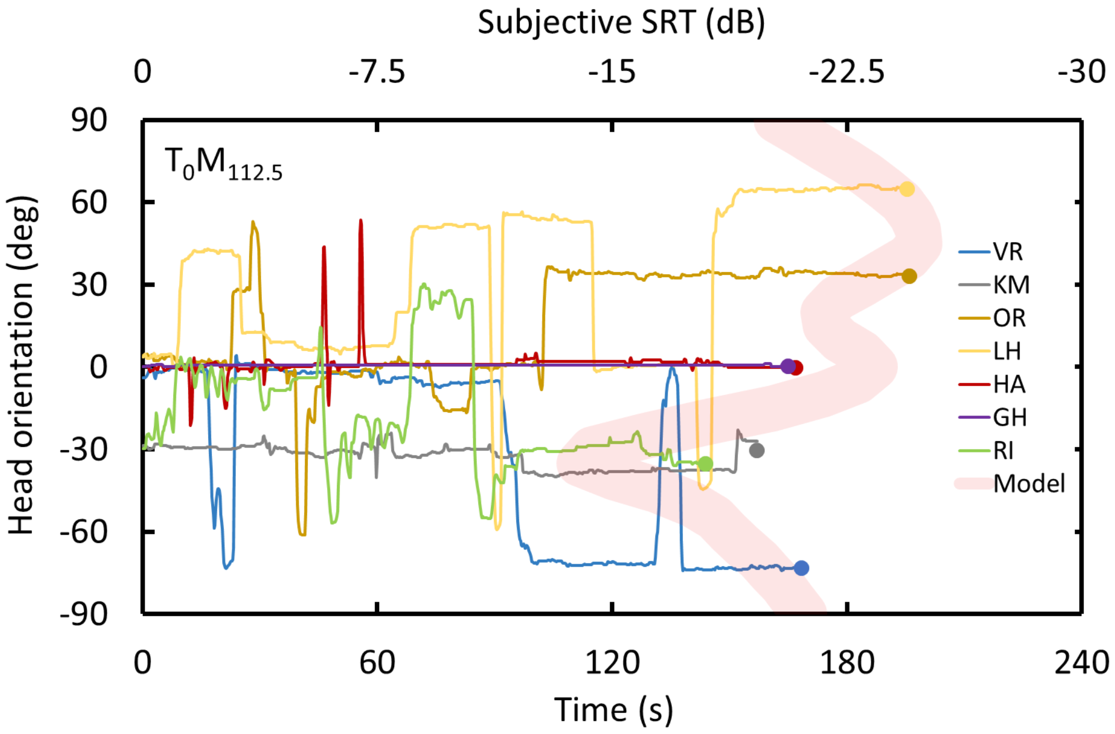


Figure 3.5: $T_0M_{112.5}$ head-orientation example tracks (labels are participant codes) against absolute clip time (lines) and subjective SRTs (filled circles) pre-instruction

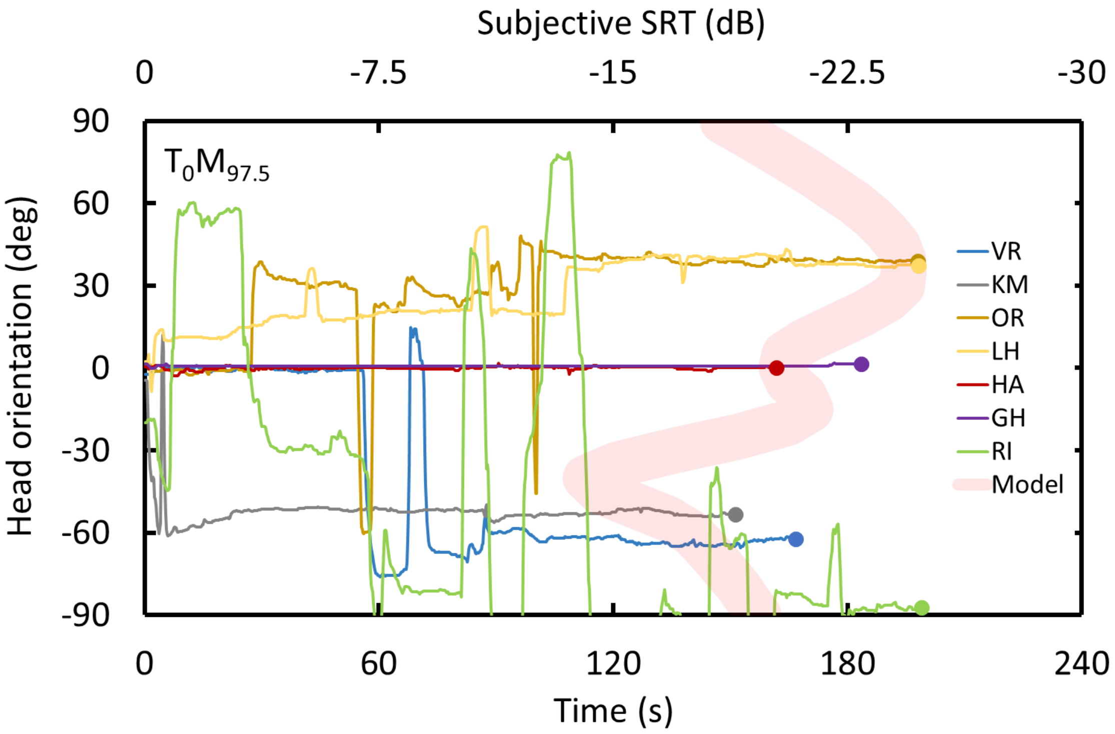


Figure 3.6: $T_0M_{97.5}$ head-orientation example tracks (labels are participant codes) against absolute clip time (lines) and subjective SRTs (filled circles) pre-instruction

3.4.2. Directed task head movements:

Only a subset of the listeners were tested post-instruction, when time allowed (9 of 16). As a result, conditions were not fully rotated across participants. What could be noted is that in the T_0M_{180} configuration (Figures 3.7), as a result of instruction, those participants who previously made little use of head orientation were able to reach much larger SNRs through exploiting head orientation. Comparison of Figures 3.3 and 3.7 shows that participants GH, HA and VR on average improved their intelligibility of speech by 8.4 dB, the average improvement across all such listeners being 6.6 dB. This is illustrated in Figure 3.8. Having made use of the same speech material in both pre- and post-instruction tasks, some of the gain found may be due to a learning effect. This could come either from acclimatisation to the speaker’s voice or from being exposed to the same material a second time. However, since the average post-instruction improvement of 6.6 dB is large, we felt that the bulk of the post-instruction improvement must stem from exploiting the benefit of head-orientation. This is further explored in the Discussion section and in Chapter 4.

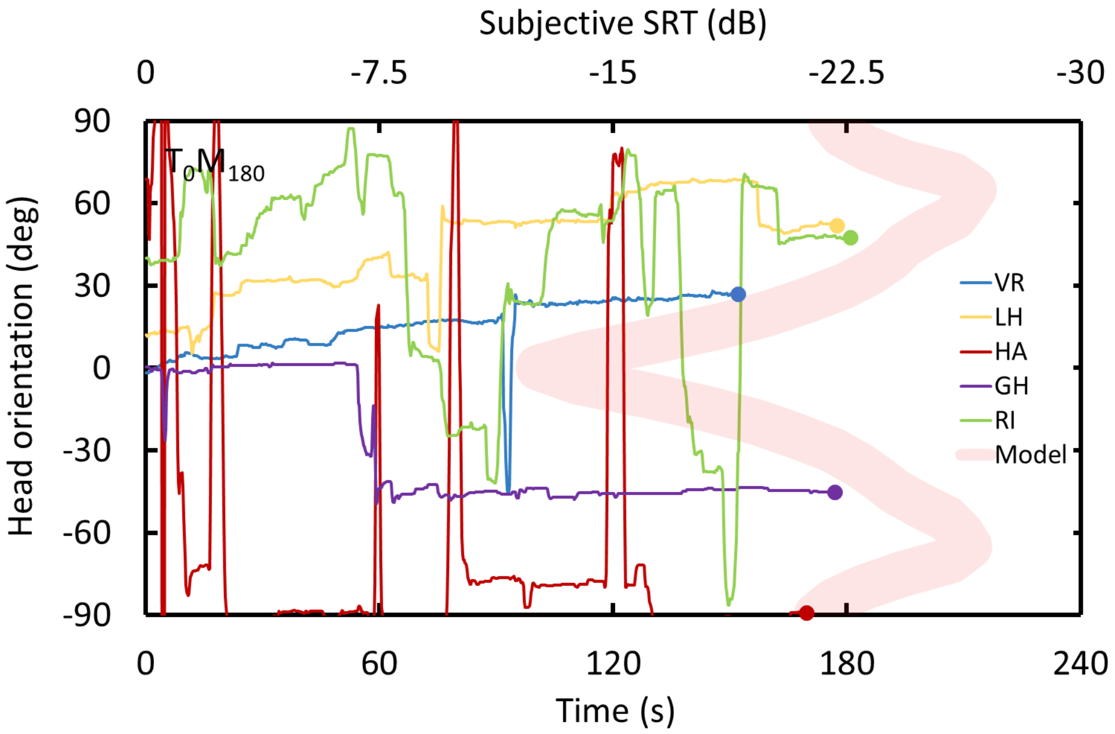


Figure 3.7: T_0M_{180} head-orientation example tracks (labels are participant codes) against absolute clip time (lines) and subjective SRTs (filled circles) post-instruction

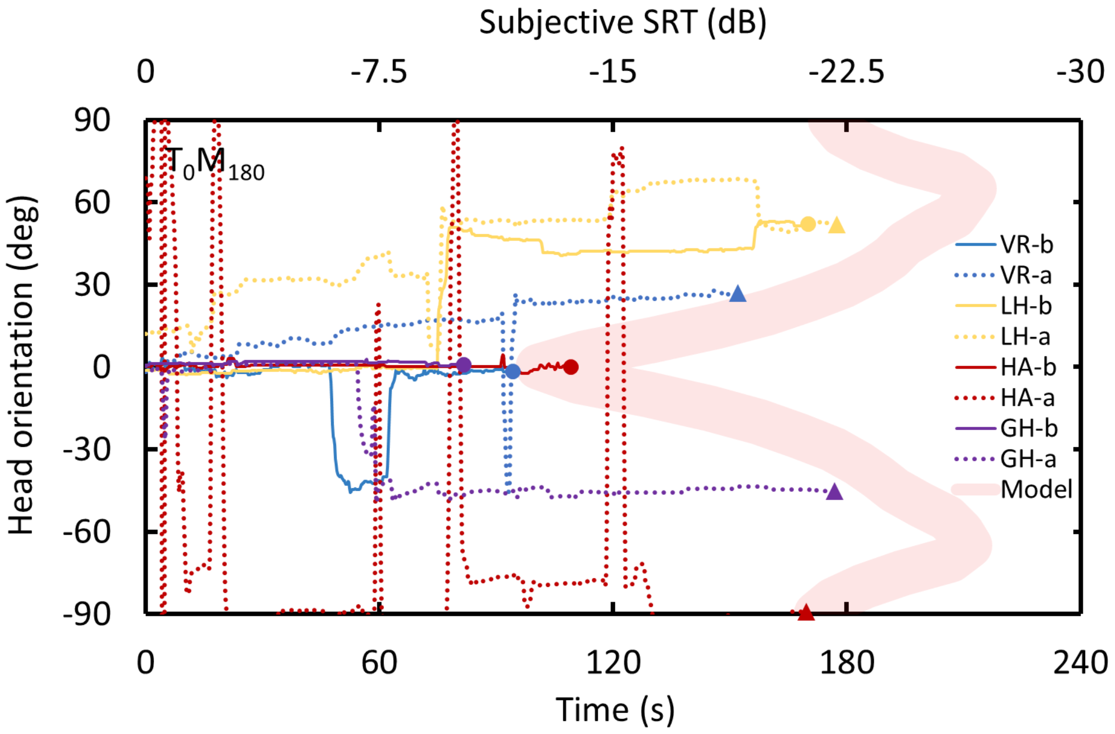


Figure 3.8: Some T_0M_{180} head-orientation tracks and subjective SRTs before (labels are participant codes, suffix b: solid lines & circles) and after instruction (label suffix a: dotted lines & triangles)

Post-instruction examples of head orientation tracks from the three asymmetrical configurations are displayed in Figures 3.9-11. These also broadly follow the model predictions. We could not clearly establish whether listeners opted for maximising SNR or target level at the better ear. After instruction, those who had not previously moved their heads improved on average by 3.6 dB. Others improved on their initial performance but some either did not exploit head orientation effectively or even persisted in reaching worse SNRs than they would have by remaining still. An ANOVA operated across all four spatial configurations and comparing pre- and post-instruction head movements (as defined above) revealed a significant effect of instruction ($F(1,8) = 6.30$, $p = 0.036$). As a result of instruction and ensuing increased head movements, subjective SRTs across all configurations significantly improved by 2.54 dB on average ($F(1,8) = 20.63$, $p < 0.002$) and remained configuration-dependant ($F(3,24) = 6.76$, $p < 0.002$), with no (instruction x configuration) interaction. Whilst listeners were generally poor at effectively exploiting head orientation pre-instruction, an immediate benefit to intelligibility in noise was thus observed from this simple instruction to experiment with head orientation, but listeners still did not exploit HOB optimally. Any overall effect of configuration on head movements disappeared as a result of instruction ($F(3,24) = 0.43$, $p > 0.7$), as expected.

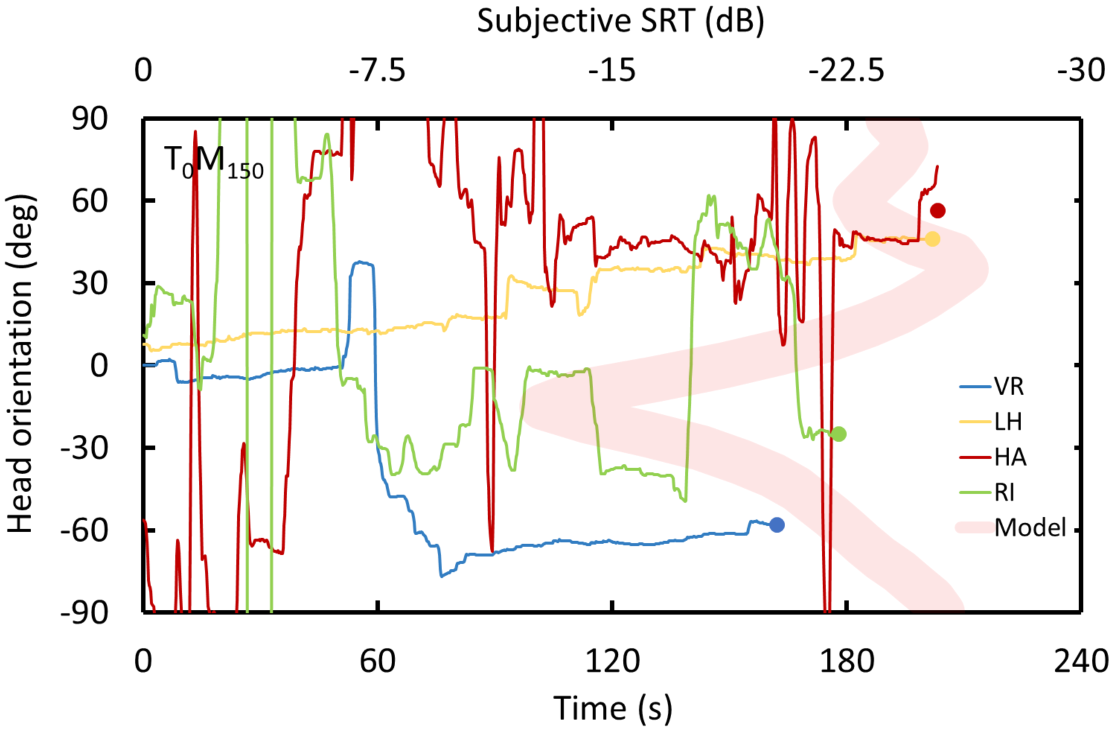


Figure 3.9: T_0M_{150} head-orientation example tracks (labels are participant codes) against absolute clip time (lines) and subjective SRTs (filled circles) post-instruction

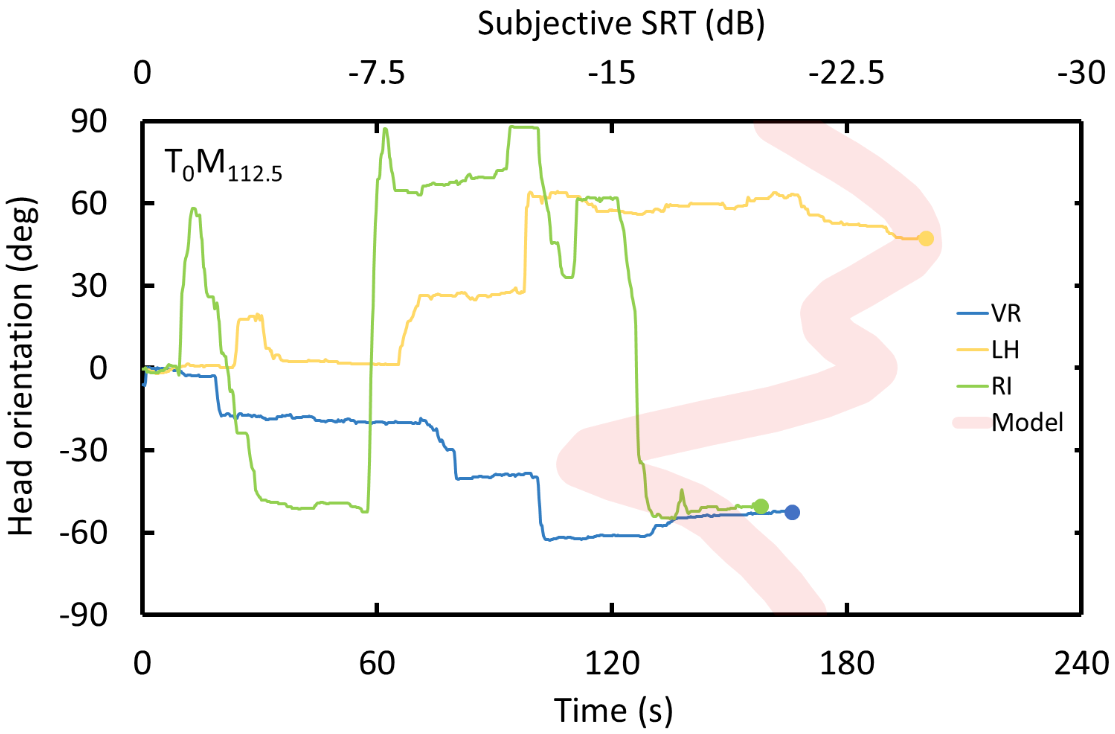


Figure 3.10: $T_0M_{112.5}$ head-orientation example tracks (labels are participant codes) against absolute clip time (lines) and subjective SRTs (filled circles) post-instruction

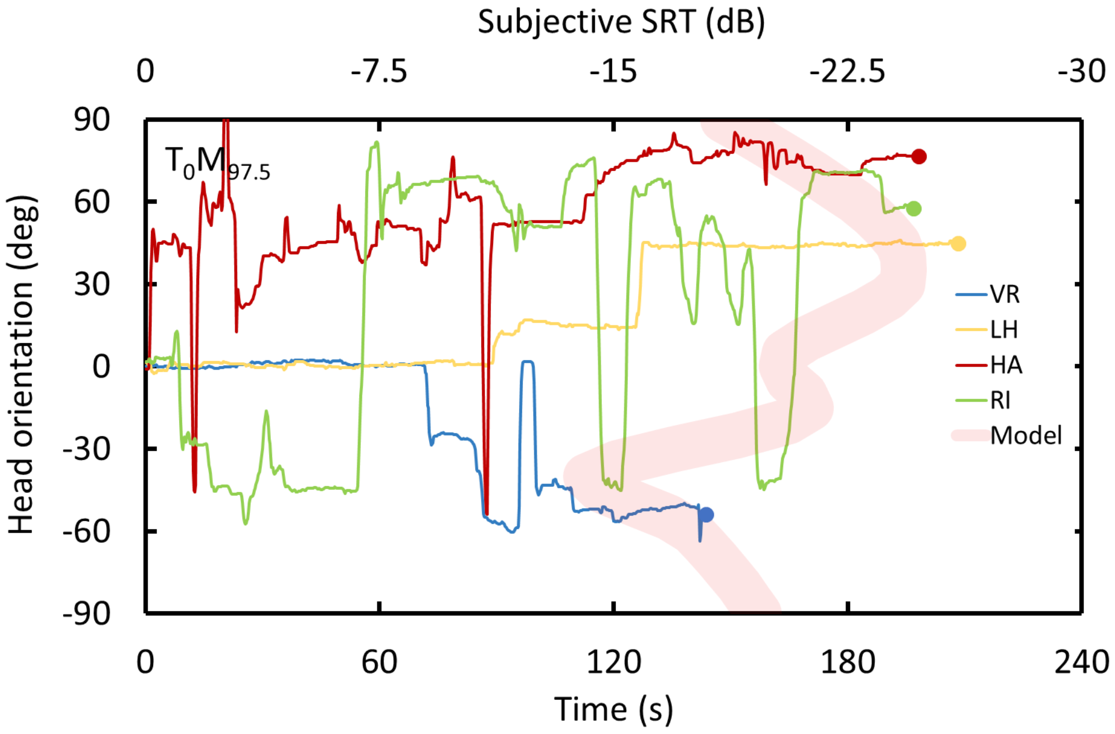


Figure 3.11: $T_0M_{97.5}$ head-orientation example tracks (labels are participant codes) against absolute clip time (lines) and subjective SRTs (filled circles) post-instruction

3.4.3. SRTs

Objective SRT measurements compiled across the first 8 participants revealed that they did not match well with predictions based on HRIRs from anechoic recordings at MIT (Gardner & Martin 1995). These anechoic predictions we hereafter call HRIR predictions. Figure 3.12 illustrates the mismatch between HRIR predictions (blue dotted line) and SRT data (orange circles) for the T_0M_{180} spatial configuration. The other three configurations displayed similar mismatches. Having initially expected the room's small reverberation level to have little impact on the predicted benefit, this assumption was reconsidered and measurements of binaural room impulse responses (BRIRs) were made in order to take the room reverberation into account in the model. BRIR measurements were made in three separate ways. First, with a B&K Head and Torso Simulator (HATS) placed on the swivel chair, making use of the chair's swivel to have the HATS face between -90°and +90° azimuth every 7.5°. Second, with a KEMAR, scanning the same azimuths in the same way. Third, with the KEMAR but keeping the chair and torso facing zero degree azimuth and rotating only the head.

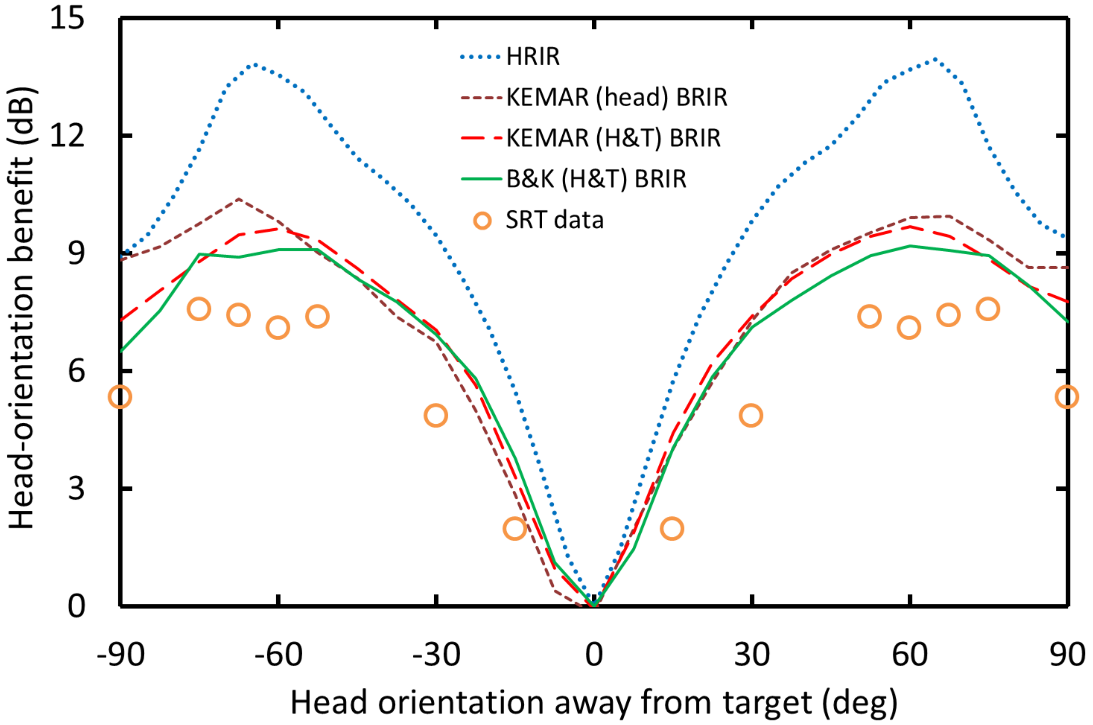


Figure 3.12: T_0M_{180} HRIR predictions, BRIR predictions and objective SRT data. B&K and KEMAR head & torso rotated together (H&T); KEMAR head rotated alone (head)

The aims were to determine whether inclusion of the room's acoustics improved the SRT fit to model predictions, to determine which manikin would make the model best fit the SRT data and to establish whether rotating head and torso together rather than the head alone made an appreciable difference in predictions.

3.4.4. BRIR predictions vs. SRTs

The BRIR predictions are also shown in Figure 3.12 (dashed and solid lines) for the T_0M_{180} spatial configuration, where they can be compared to the HRIR predictions and the SRT data. Head rotation benefit is calculated by taking the difference between SRTs measured or predicted with a given head orientation and the SRT measured or predicted when facing the front. Despite the use of a sound-deadened room, there was a significant predicted effect of reverberation on HOB. Maximum benefit was predicted to drop from 14 dB to 9 dB. We concluded that even a modest level of reverberation (RT60 = 60 ms) must be responsible for the effect seen. As discussed in Culling et al. (2012), this illustrates how RT60 is not a reliable measure of the impact of reverberation. All three sets of BRIR predictions were very similar. However, when comparing head and torso rotation to head rotation alone with the KEMAR, divergence between the two data sets only occurred beyond +/-60°. Given that any head rotation with respect to torso beyond 60° would be unnatural when attending to speech, we concluded that the effect of the torso position with respect to the head was not relevant in the context of this study. The other three spatial configurations confirmed the findings above. Figure 3.12 also displays objective SRTs averaged across the first 8 participants.

Figures 3.13a-c show the best model fit to the observed data (obtained with the B&K BRIR predictions). The SRT data was moved along the HOB axis so as to equalise the means of predictions and data across the four spatial configurations. The disparity between data and predictions was typically within less than 1.2 dB (RMS error = 0.77 dB), with the standard error of the means (SE) of the SRT data not exceeding 0.56 dB and averaging 0.41 dB. The only exceptions to data fitting the model within 1.2 dB were found where the sharpest slope in benefit per degree of head rotation was predicted. There, the discrepancy was 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 prediction. It proved difficult for listeners to maintain a fixed and correct head orientation whilst focusing on the listening task at hand. It is plausible that listeners may have deviated at times by as much as 5° from the correct head orientation, which accounts for the largest deviations from prediction. This constitutes a justification for the SRT data offset operated in Figure 3.13 and is further validated in Chapter 4.

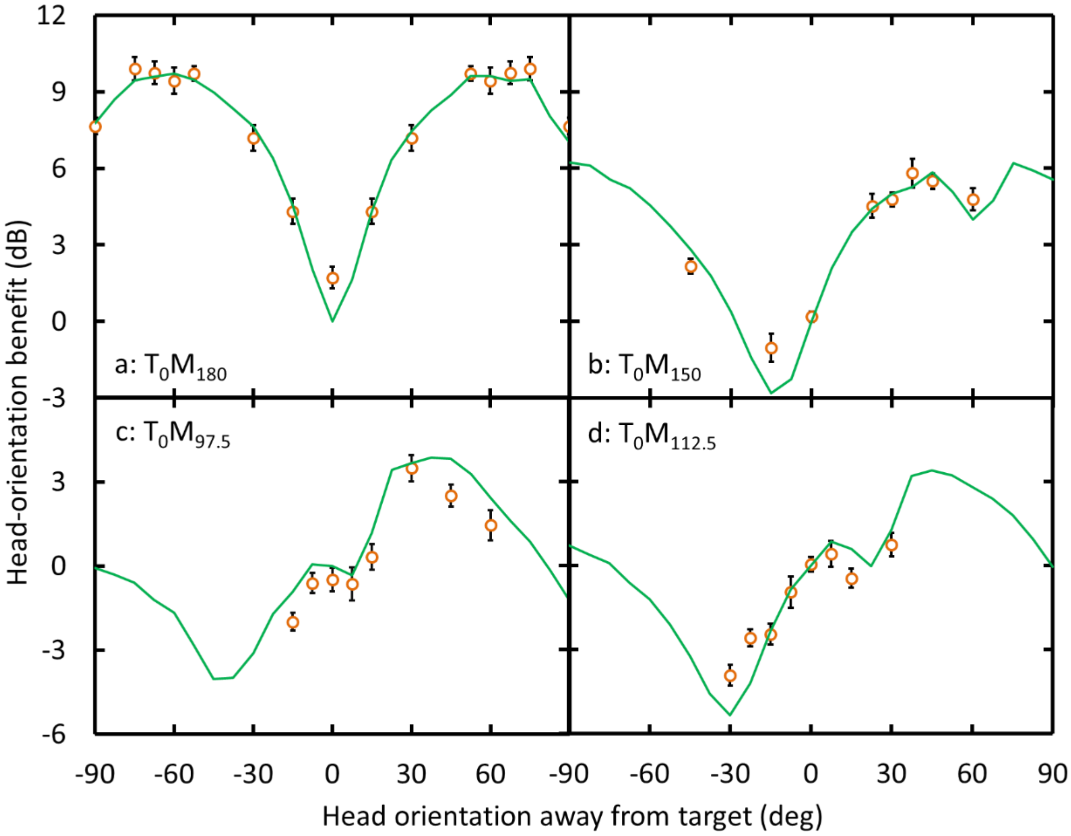


Figure 3.13: Predicted HOB (lines) and objective SRTs (circles) averaged over 16 participants for all spatial configurations. Error bars are standard error of means

Operating an ANOVA in each spatial configuration, a significant effect of head orientation on HOB was found ($F(6,90) > 15.4$, $p < 0.001$). Within each configuration, all pairwise comparisons reveal significant HOB differences where the error bars (SE) do not overlap in Figure 3.13. A comparison of speech-facing SRTs reveals, 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 configuration now broadly show the same trend. The clear definition of minima and maxima close to the speech-facing orientation found in HRIR predictions in Figure 3.2 is very much dampened. Asymmetrical configuration prediction curves have also been shifted to the right by about 10 degrees. As a result all three asymmetrical configuration now generally show a detriment of head turns to the left and a benefit to the right. The subtleties involved in the original choice of asymmetrical spatial configurations are therefore irrelevant.

3.4.5. Confidence in the predicted effect of reverberation

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 3.14 compares predictions for both manikins' head and torso rotated together, with and without room reverberation. A good match was found between MIT HRIR predictions and our anechoic KEMAR condition (RMS error = 0.5 dB) for each spatial configuration.

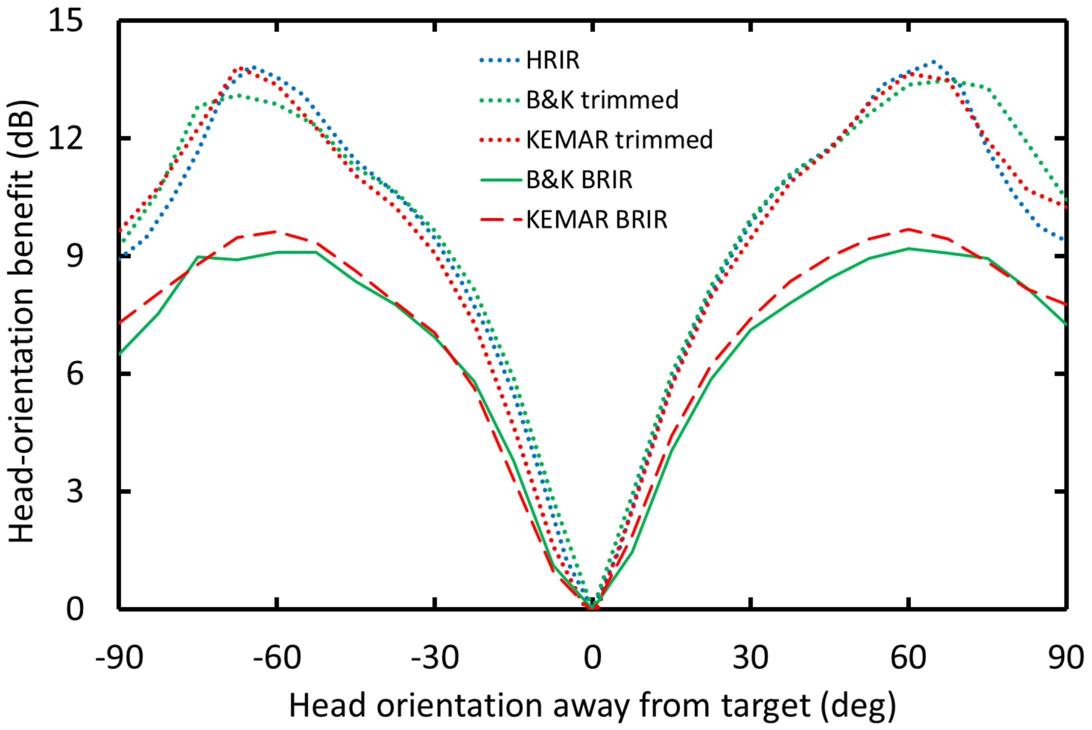


Figure 3.14: T_0M_{180} HRIR (blue dotted line), BRIR predictions (solid & dashed lines) and predictions from BRIRs trimmed to direct sound (dotted green and red lines) with KEMAR and B&K manikins, head & torso rotated together

3.4.6. Predicted impact of reverberation for cochlear implant users

Since the original motivation for this pilot experiment was to provide a NH baseline for further work with bilateral and unilateral CI users, it was felt important to compare how much room reverberation was predicted to impact on both types of listeners' HOB. In order to assess the effect we used the model to compare predictions for NH listener and CI users. For the bilateral-CI predictions, the same assumptions as in Culling et al. (2012) were made that both implants performed equally well and that CI users do not benefit appreciably from binaural unmasking. Figure 3.15 shows HRIR and BRIR predictions for the T_0M_{180} configuration for NH & CI users. CI user predictions were derived by only considering the BE component of SRM. Thus the difference between NH listeners' and CI users' predictions in Figure 3.15 represents the predicted effect of BU (see Chapter 2). The effect of moderate reverberation on HOB was calculated by subtracting HRIR predictions from BRIR predictions and is plotted in Figure 3.16. Here, reverberation is predicted to affect CI users' HOB half as much as NH listeners'. This can be explained by reverberation affecting mostly the BU and proportionately much less the BE contribution to NH SRM. More than half of the BU effect in NH listeners is removed by reverberation. This is represented by the gap between NH and CI user predictions in Figure 3.16. The net result is that there is much less predicted difference in HOB between NH listeners and CI users in this moderately reverberant room than in anechoic conditions. This observation is expected to extend to more reverberant conditions as they would further erode NH listeners' BU, along with their BE benefit.

Close analysis of the BRIRs showed that the largest reflections are the first reflections from the floor and the ceiling (typically 10-15 dB below the direct sound). The third largest reflection is from the wall opposite to the sound source, but since the walls were sound-treated, this reflection is 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 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°, most common when attending to speech. Indeed, as the first two reflections come from the same azimuths as the direct sound, they should not affect ITDs and ILDs in a manner detrimental to SRM because they remain coherent with the direct sound, at moderate head angles. The lateral reflections, although weaker, directly affect ITDs and ILDs. The blurring of the ITDs they cause reduces the normally hearing listeners' ability to exploit binaural unmasking whilst a small change in ILDs affects the head-shadow effect in all listeners much less.

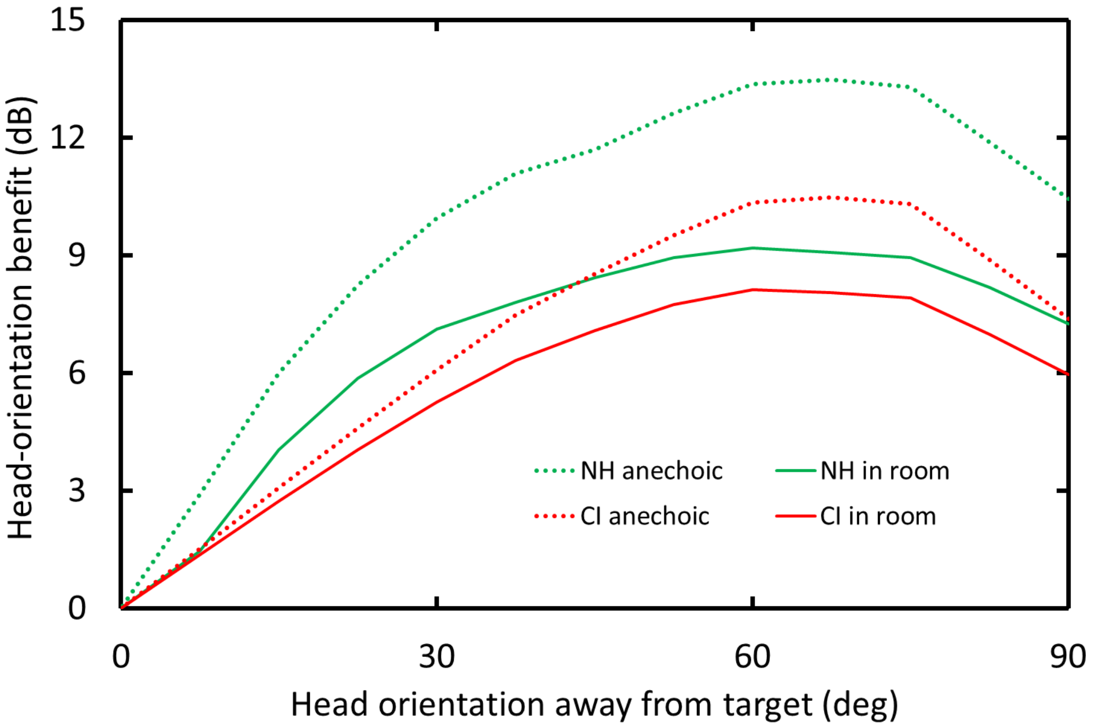


Figure 3.15: T_0M_{180} B&K BRIR predictions (solid lines) and predictions from BRIRs trimmed to direct sound (dotted lines) for NH listeners and CI users

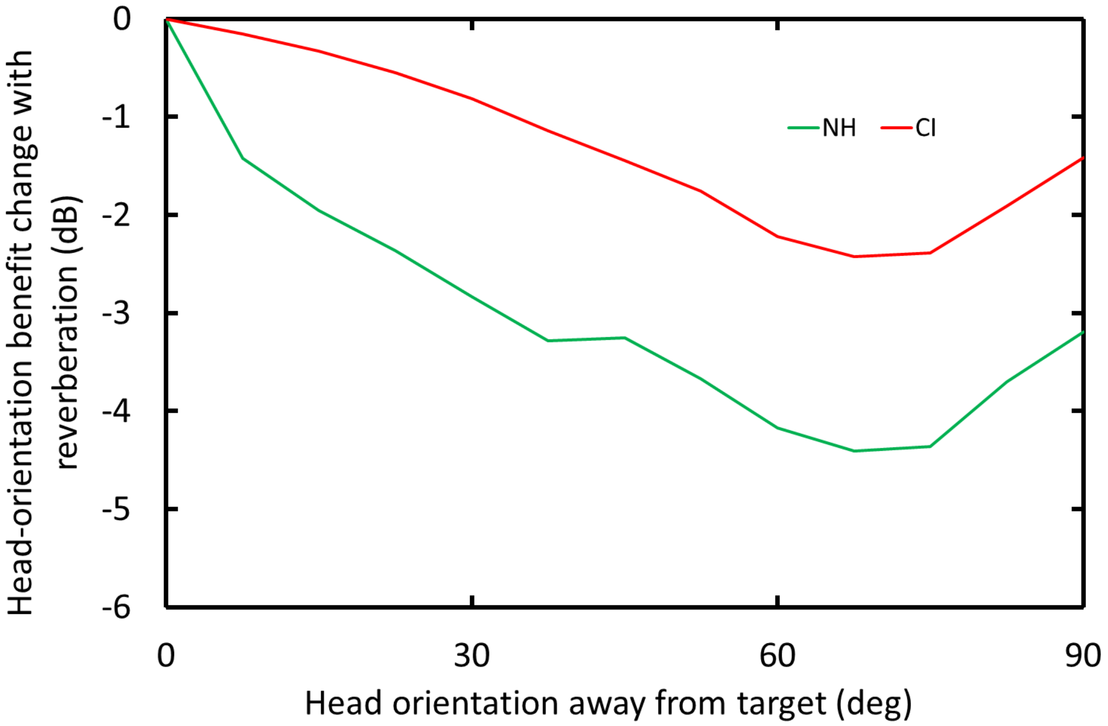


Figure 3.16: Predicted change in T_0M_{180} head-orientation benefit due to moderate reverberation for NH listeners and CI users

3.4.7. Subjective SRTs versus BRIR predictions

The directed version of the free-head task was performed whenever time allowed and particularly when listeners did not move their head in the undirected paradigm. Combining data pre- and post-instruction, subjective SRTs are plotted in Figure 3.17 against BRIR HOB predictions for the settling head orientations. The latter were defined as head orientations averaged over the last 10 seconds of a head track. A linear regression applied to each plot resulted in a correlation coefficient ranging from $r = 0.37$ to 0.71 (slopes ranging from 0.51 to 1.08) and correlations were found significant only for T_0M_{180} ($p < 0.001$) and $T_0M_{97.5}$ ($p < 0.005$). Three data points were removed when listeners brought their hands to their ear to aid their intelligibility of speech, thereby going against instructions.

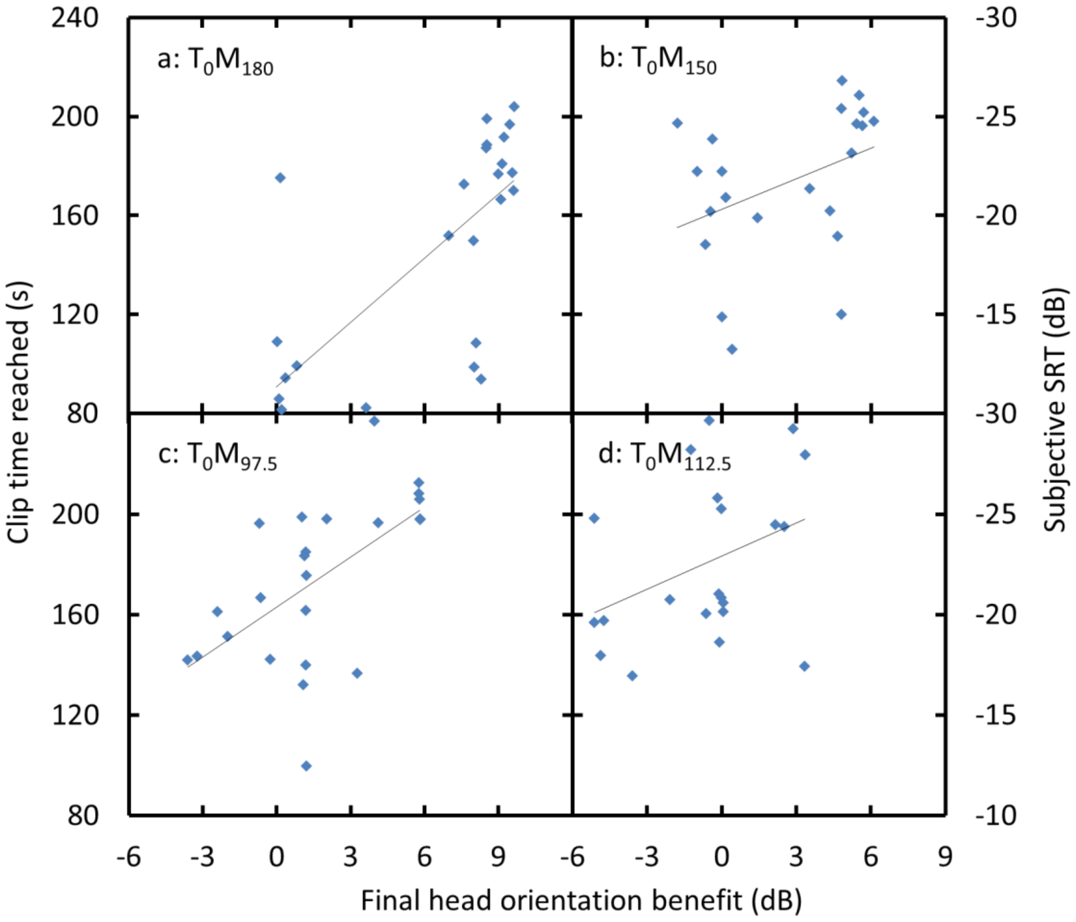


Figure 3.17: correlation between subjective SRTs and head-orientation benefit predicted for the final head orientation in each spatial configuration

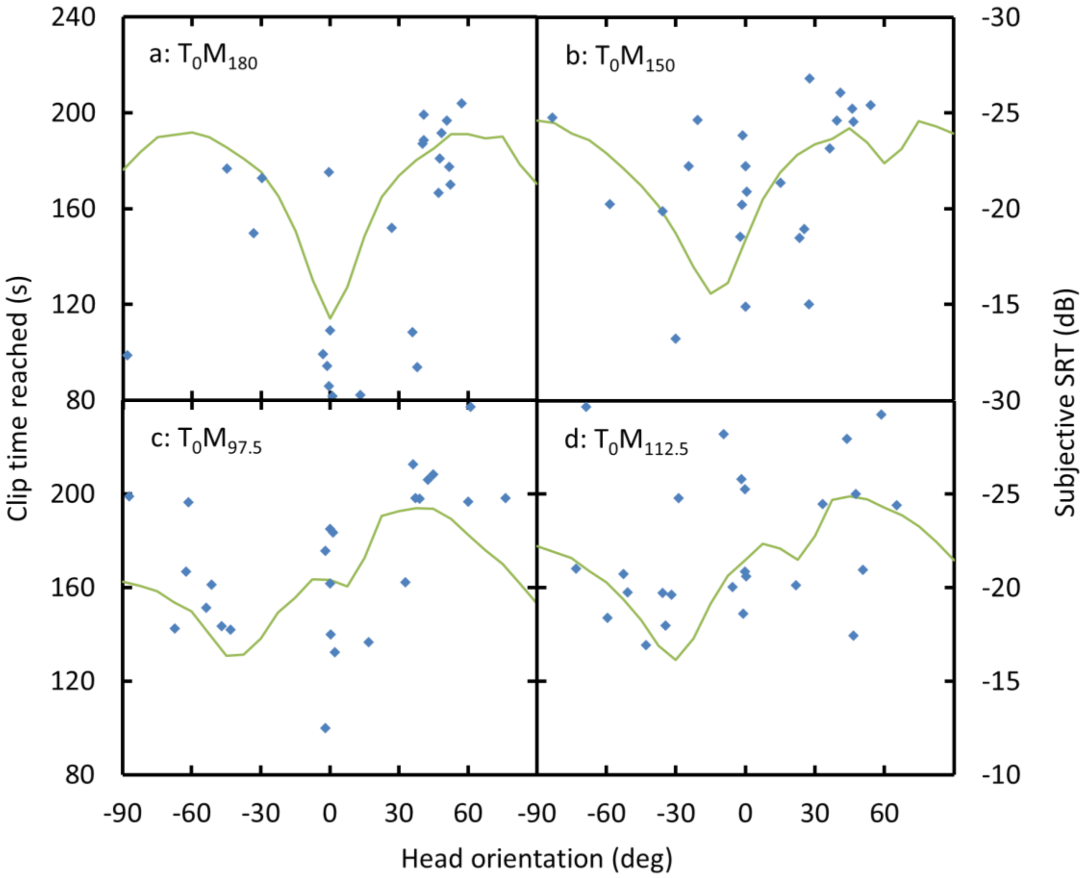


Figure 3.18: Subjective SRTs in the context of model predictions for each spatial configuration

Figure 3.18 plots, for each spatial configuration, predictions and subjective SRTs against final head orientation. The model predictions were offset along the ordinate through equalisation of the BRIR predictions and data means across all four configurations. The data very loosely followed the predictions since they contained individual variations. The noise is attributed to a combination of variability in objective SRTs found across listeners and variability in the criteria used by listeners to judge that they had lost track of the meaning of the clips.

In the T_0M_{180} configuration, over the trials where participants moved their heads, they turned to the right twice as often as to the left when they experienced 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 again turned to the right twice as often as to the left. In the other two asymmetric configurations the numbers turning to the right and to the left were approximately equal. Overall, when benefit of rotation is symmetrical, there seems to be a bias towards presenting the left ear to the target and when the benefit is asymmetrical, there is no clear bias anymore.

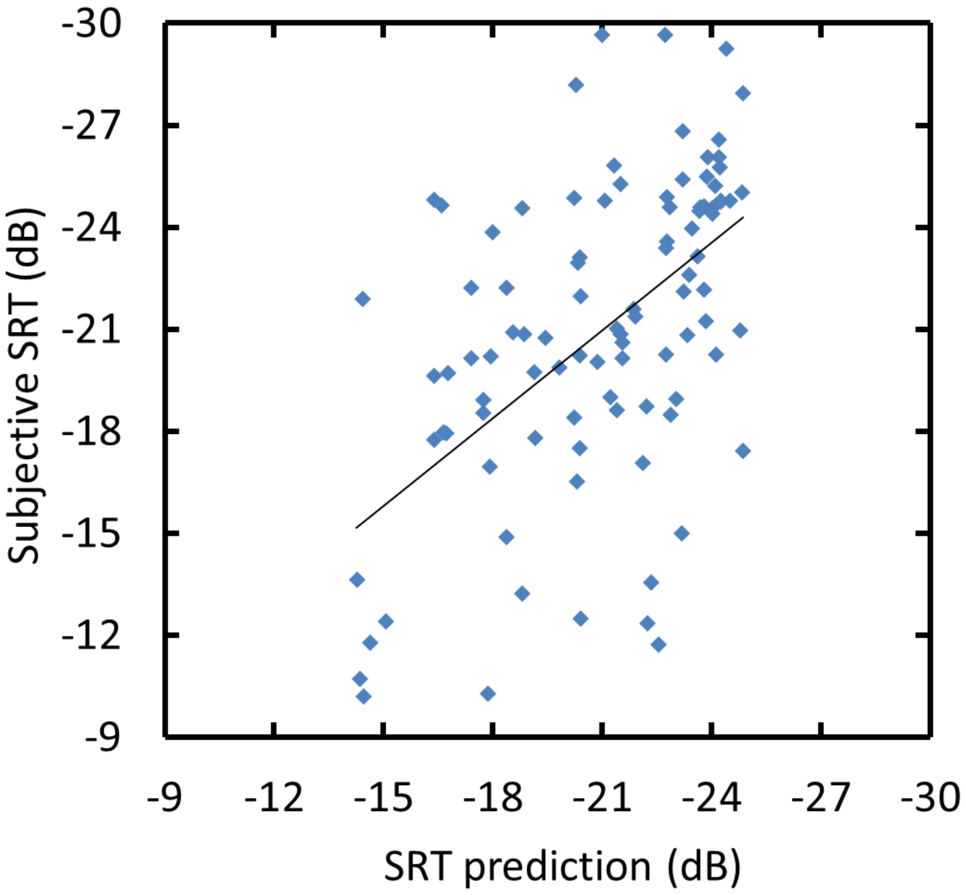


Figure 3.19: Correlation between subjective SRTs and SRTs predicted for final head orientations

The four plots of Figure 3.17 were combined in Figure 3.19, with subjective SRTs now plotted against SRT predictions for the corresponding final head orientations. The SRT predictions were offset so as to equalise predictions and data means. A significant correlation was found between the two (r = 0.54, t(96) = 6.31, $p < 0.001$). This confirms that the more listeners exploited head orientation the more they could understand of the clip. The slope of the regression line was 0.86 (0.13 SE), indicating somewhat less benefit of head orientation than predicted.

3.5. 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. The SRT data matched the HOB predictions within typically less than 1 dB over an 8 dB range of benefit. We found that feeding the model with BRIRs acquired in the listening room rather than HRIRs was critical to a good match. Indeed reverberation was responsible for a reduction in predicted benefit for NH listeners from 14 dB to 9 dB at the optimum head orientation and in the symmetrical spatial configuration. According to the model, CI users should be affected by reverberation less than NH listeners in this room, as its moderate reverberation reduces normally hearing listeners' binaural unmasking much more than it reduces better ear listening. This is assuming CI users typically do not benefit from binaural unmasking and is expected to extend to more reverberant situations.

When attending to a diminishing speech level in a fixed noise level and in the same four spatial configurations, 60% of listeners in an undirected paradigm were found to make use of head orientation to improve their speech intelligibility. However it is clear that only half of them made use of an effective strategy. Few listeners seem to have learned how to make use of sound localisation effectively to optimize their strategy. In the T_0M_{150} configuration, where the largest HOB slope was found around the speech direction (benefit to the right and detriment to the left), in nearly twice as many trials, listeners turned their heads to the right (11) rather than to the left (6). However the numbers were almost equal in the other two asymmetric configurations, even though turning one's head to the left was detrimental. Considering our initial hypotheses on free-head strategies, the asymmetrical spatial configurations do not clearly show that participants' strategies were influenced by the direction of the SRM slope. Moreover, in the symmetrical configuration, listeners preferentially turned to the right over more than two thirds of the trials in which they made use of head orientation, suggesting a general bias. Our data presents only weak evidence that sensing the SRM slope whilst scanning for intelligibility improvement motivates participants to rotate their head the correct way. The T_0M_{150} data, however, suggests that a steeper slope may help listeners make the correct decision.

We found that across all four configurations the strategies adopted were extremely varied. Some listeners did not move and therefore demonstrated no strategy other than listening hard to the speech. They lost track of the speech earlier than other listeners as a result. 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 were the most interesting in that they seemed to have a much more developed strategy which allowed them to move straight to the optimum head orientation without the need for scanning. This may be evidence that those listeners make use of localisation of the sound sources and work out, from experience, where to position their heads before they move. Our best performing participant, OR, who excelled in all tasks, had extensive experience of working in a recording studio as a musician. This may have led him to use sound source localisation to deduce the optimal head orientation and home in on a target amongst competing sources, e.g. a specific instrument somewhere in the studio.

The model shows that a correct strategy is typically to point one's head towards 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. This was clearly a design flaw, in retrospect, since masker position could have been counterbalanced across participants by testing in mirror configurations. It is therefore not possible to distinguish between a natural response to the asymmetry of cues and a plausible, general bias towards turning one's head one way rather than the other, thereby favouring a particular ear. 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 worsened. Some of them, even when given a second chance in the directed paradigm, persisted with this erroneous strategy. This tells us that they had achieved some degree of localisation of the noise source but had failed to exploit that information correctly. Hence a lot can be done to help people optimize their head orientation strategy when attending to speech in noise. The question of whether this translates to CI users is studied in Chapter 5.

Brimijoin et al. found that asymmetric hearing-impaired listeners favoured maximizing signal level over SNR at their better ear (Brimijoin et al. 2012). Unfortunately, our choice of spatial configurations and lack of statistical power (too few trials) did not allow us to establish whether our listeners favoured the same strategy. The original anechoic predictions were somewhat misleading as revised predictions show that for each configuration, the head orientation leading to maximum SRM is close to the 60° 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°. They may well have been.

Young normally hearing listeners were poor at making effective use of the cues available to them since in a third to half of the asymmetric trials, those who moved turned the wrong way and a third of the listeners did not spontaneously move at all. 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 realising it, made use of head orientation. Older normally hearing listeners, whose SRM is known to be reduced (Dubno etal., 1984; Glyde etal., 2011; Helfer etal., 2010; Marrone etal, 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. Both CI users and older, normally hearing adults were tested for this and results are reported in Chapter 5.

In their study, Brimijoin et al. had chosen to combine the free-head orientation task with a short adaptive task that initially brought the SNR level close to the listeners' SRT when facing the speech (Brimijoin et al. 2012). This was intended to motivate listeners to make use of head orientation. Their use of short-sentence presentations throughout a run may instead have resulted in the authors not obtaining any significant levels of spontaneous head movements in normally hearing listeners (pers. comm.). Our findings contrast with Brimijoin et al.'s. This suggests that our free-head task's gradual SNR diminution, made possible by keeping objective SRT measurements separate, provided listeners with a more effective motivation to move their heads.

The SRT experiment showed that the model was accurate at predicting head orientation SRM. The listeners who were able to make effective use of the cues available to them could either spontaneously excel at the free-head task or make significant progress post-instruction. This seemed encouraging in terms of training, should it turn out to benefit CI users. Later testing of CI users, described in Chapter 5 and performed with similar paradigms, will in part aim to find out how practical and useful such training could be.

3.6. Conclusion

The pilot experiment presented in this chapter further validated the Jelfs et al. (2011) model of SRM for predictions of head-orientation benefits. 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 were relatively poor at making effective use of cues available to them in order to orient their heads optimally. In 40% 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 or erratically. They generally performed poorly, sometimes going as far as choosing head orientations that made speech intelligibility worse even though they might have passed through beneficial head orientations and/or localised the masker position. Repeating the task after instruction, listeners generally improved. This indicated that training on how to optimise one’s head orientation strategy could be beneficial, practical and fast.

4. AUDIO & AUDIO-VISUAL BASELINE WITH YOUNG NORMALLY HEARING ADULTS

In Chapter 3, we saw how accurately the Jelfs at al. model of SRM could predict changes in the intelligibility of speech in noise for young, normally hearing adults as a function of head orientation away from the target speech. We observed, for a variety of masker spatial separations, how poorly listeners exploited head orientation when attending to speech in audio-only presentation with gradually diminishing signal-to-noise ratio. Following instruction to make use of head orientation, listeners were found to improve, although some persisted with poor performance. Since our core interest was in discovering whether CI users can benefit from exploiting head orientation, the next natural step was to develop paradigms tailored for CI users. As in the preliminary study reported in Chapter 3, the pilot experiments presented here were divided into two sets of runs: free-head, behavioural runs and fixed-head, SRT-measurement runs, the results from one set informing those of the other. In order to make testing more relevant to the everyday situations that listeners may face when listening to speech in noise, testing was also performed with AV presentations. AV conditions are important to CI users, because hearing-impaired people are well known to be more reliant on lip-reading (e.g. Mitchell & Maslin 2007; Giraud et al. 2001). We therefore set out to measure LRB for speech intelligibility in noise, assess the impact of visual cues on the listeners' behaviour and determine whether head orientation away from facing the speech could be part of a more effective listening strategy for CI users. Initially, the experiment had to be piloted with young NH adults. Chapter 4 presents the changes made to paradigms and speech material in order to accommodate CI users, then details outcomes for young NH adults.

4.1. Spatial configurations and head orientations

Informed by the preliminary study reported in Chapter 3 and given that most published studies tested SRM in CI users with speech in front and masker at 90 º, three spatial configurations and two head orientations were selected. The chosen spatial configurations were with target and masker collocated at 0º versus target at 0º and masker at 90 º or 180º. The head orientations for the objective SRT runs were either facing the speech or with the head rotated by 30º. A 30º head turn was chosen to demonstrate a head-orientation benefit that we anticipated may not have a detrimental impact on lip-reading. Although gaze can be maintained up to 45º, it was felt that 30º would be sufficient to generate a strong HOB, whilst still permitting a realistic and comfortable listening posture. Moreover, most of our CI-user volunteers were going to be older adults and likely to wear glasses, which would probably rule out maintaining a gaze direction greater than 30º. As a HOB can only be attained when sound sources are spatially separated, the collocated situation was not tested with a 30º head turn.

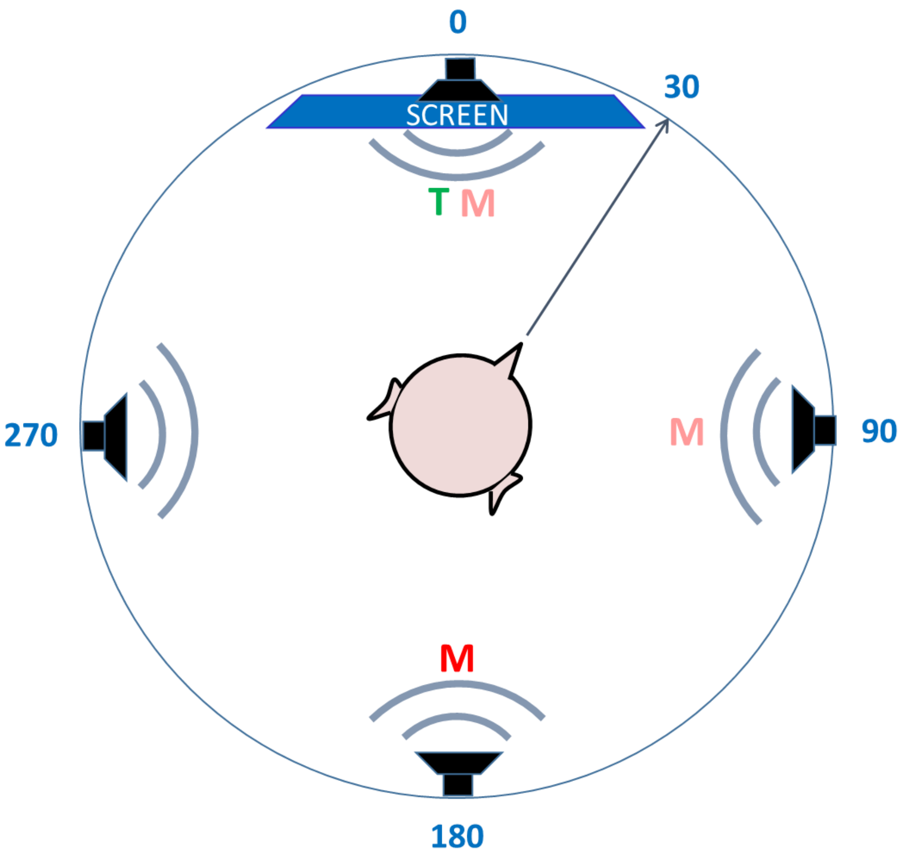


Figure 4.1: Schematic of the SRT spatial configurations. Codes incorporate masker separation and head orientation (highlighted is the $H_{30}M_{180}$ configuration)

For the SRT experiments, spatial configuration and head angle were combined for simplicity in a new configuration code such as $H_{30}M_{180}$, the configuration highlighted in Figure 4.1, where H_{30} indicates that the head is rotated by 30º from the target speech and M_{180} that the masker is separated by 180º from the target speech.

Figure 4.2 plots normally hearing model predictions derived from BRIRs acquired in the test room. In the T_0M_{180} configuration, a HOB was predicted equally for orienting the head to the left or to the right of the target. In the T_0M_{90} configuration, turning to the left leads to a positive HOB prediction whilst turning to the right leads to a loss in intelligibility. The expectation from model predictions was that a favourable 30º head turn would either provide the bulk of the attainable audio SRM in the T_0M_{180} configuration (at $H_{30}M_{180}$) or provide the maximum attainable SRM in the T_0M_{90} configuration (at $H_{30}M_{90}$), as highlighted by the red dots in Figure 4.2. The H_0M_{90} configuration was chosen in order to allow us to compare our new audio-only data with prior studies (e.g. Culling et al. 2012). The T_0M_{180} configuration was selected to maximize benefit of head rotation according to the model predictions (see Chapter 3). The collocated configurations H_0M_0 and T_0M_0 acted as reference for all SRT measurements and free-head tracks respectively.

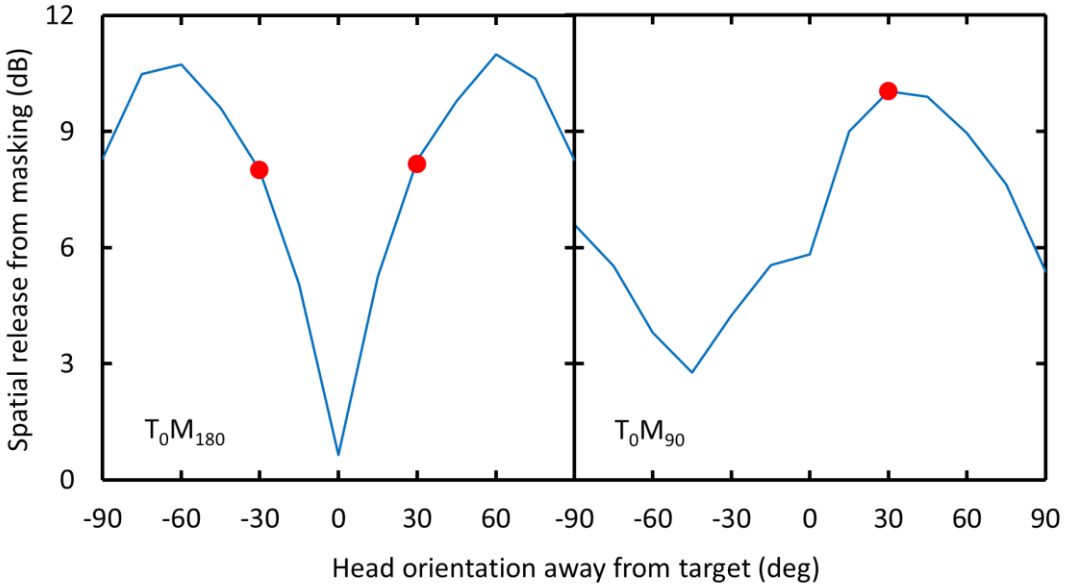


Figure 4.2: Model predictions in the T_0M_{180} and T_0M_{90} configurations. Red dots highlight the best predicted SRM at 30º head orientation(s)

4.2. Hypotheses

For the free-head orientation strategy experiment we expected, as found in our Chapter 3 pilot study, that NH listeners would be poor at spontaneously exploiting the head-orientation benefit and that approximately half of the listeners would probably not spontaneously move their heads. We hypothesised that AV presentation would reduce the tendency for head turns on the assumption that a more natural or socially acceptable behaviour is to face the speaker (Plomp 1986).

For the SRT experiment we expected SRM to follow the model predictions and to find a significant 30º HOB. We thought that the presence of visual cues in the AV modality would improve SRTs. We also hypothesised that the final head orientation reached and subjective SRTs in the free-head task would reflect either directly measured objective SRTs (where available) or modelled head-orientation benefit.

4.3. Materials & Methods

4.3.1. Participants

22 NH participants aged from 18 to 22, age mean 20, were recruited from the undergraduate Cardiff University population. The first 10 performed the SRT tasks whilst the last 12 performed the free-head-orientation task.

4.3.2. Laboratory setup

Since later testing was planned in two locations to enable easier access for CI users, a mobile laboratory-equipment kit was devised to allow easy transfer of testing between locations. All tests reported in this chapter were performed in the Cardiff University test room described in Chapter 3. As schematically shown in Figure 4.1, 4 Cambridge Audio Minx loudspeakers were arranged at cardinal positions around a 1.5 m radius circle centred on the listener's head, itself centred on the axis of a swivel chair positioned centrally in the room. The cardinal directions were parallel to the room's walls. The loudspeakers were driven by a 6-channel Auna solid-state amplifier, driven by an ESI MAYA-USB44+ 4-channel digital-to-analogue converter. The speakers were fitted 1.3 m above floor level so as to match the average head level of the participants, once sat on the swivel chair. A 17-inch video monitor was positioned below loudspeaker 1. Loudspeaker 1 defined the 0º orientation and was used to present the target speech. A shaving mirror was fitted to a fifth, mobile speaker stand. The mirror was used to assist listeners in adopting the correct head orientations during the SRT runs as in the preliminary SRT experiment.

All stimuli were controlled by bespoke Matlab programs. In the audio-only condition, they were presented directly from Matlab, making use of the Playrec toolbox (Humphrey 2008-2014). In the AV condition, the target was routed through the VLC player (VideoLAN). Each channel of the audio chain was found to be sufficiently well matched to the other three for our purpose, both in level (all within 0.5 dB SPL) and spectral response. This was verified by acquisition of impulse responses and subsequent comparison of excitation patterns, and by A-weighted sound-level meter measurements at the listeners' head position. A Microsoft Lifecam 5000 digital video camera was fitted on the ceiling exactly above the listener's head. The RT60 of the two sound-deadened room was measured from the impulse responses to be circa 60 ms using the reverse integration technique (Schroeder 1965).

4.3.3. Modifications to the standard SRT protocol and rationale

Three changes were made to the adaptive threshold method described in Section 3.3.4 (referred to herein as the 'standard' protocol) in an effort to better tune the test to CI users. Instead of using IEEE sentences, high predictability SPIN sentences, as devised by Kalikow et al., were selected (Kalikow et al. 1977). CI users are sometimes unable to recognise a particular word even at very high SNRs. In the high-predictability SPIN sentences, the words preceding the target word provide a context that makes predictability of the target high. This redundancy was expected to assist CI users and reduce the risks of their not being able to hear the target word. There remains a risk, however, that even with the context, a particular keyword may remain obscure. In the standard SRT procedure first developed by Plomp & Mimpen (1979) this can be problematic, because the adaptive phase of the measurement does not begin until the first sentence is partially intelligible. Normally, the first sentence would be repeated with 4 dB SNR increments until the target word is correctly identified. Instead, we limited this to a maximum of 4 presentations before presenting a new sentence at the previous presentation SNR. This way, should the first sentence key-word be particularly hard to identify for an individual, the start of the staircase would not be delayed too much. Following 4 unsuccessful repeats of the first sentence, a new sentence could be presented a maximum of 3 times before being renewed, starting at previous presentation SNR and then increasing SNR in 4 dB increments. As it turned out, CI users required no more than 2 sentences before they started the adaptive phase (see Chapter 5). Once the staircase commenced, SNR was changed in ±2 dB steps as in the protocol described in Section 3.3.4. However, each sentence was presented up to 3 times at increasing SNRs until the keyword was identified. Once again, after 3 unsuccessful repeats, a new sentence would be presented at the previous presentation SNR. Repetition of sentences following unsuccessful trials was intended both to make the task easier for CI users and to allow more efficient use of the relatively small number of SPIN sentences.

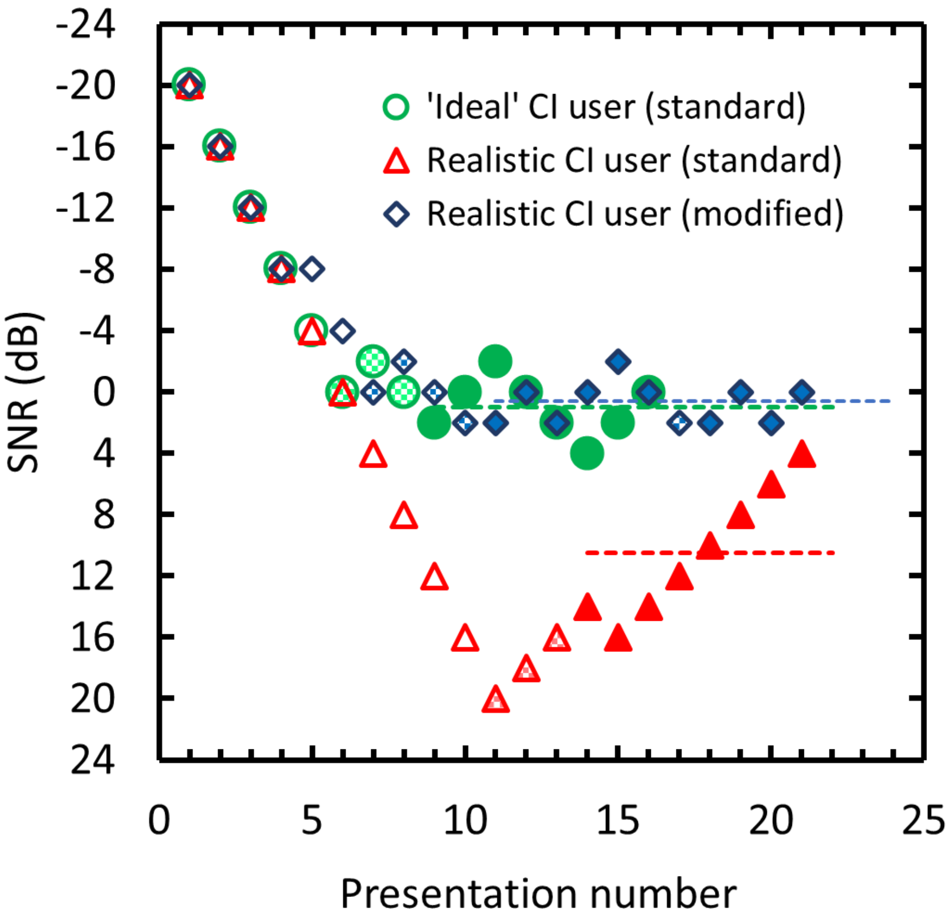


Figure 4.3: Hypothetical staircases for an 'ideal' CI user in the standard protocol (circles) and a realistic CI user in the standard (triangles) and modified (diamonds) protocols. Pre-adaptive phase SNRs (open symbols), discarded SNRs (textured symbols) and SNRs (filled symbols) averaged to compute the SRT (dotted lines)

Figure 4.3 shows hypothetical CI-user adaptive tracks in the standard and modified protocols and illustrates why we felt that modifying the protocol would help. First, an 'ideal' CI-user adaptive track in the standard protocol (green circles) shows how a CI user might perform if they had no difficulty in recognising any of the target words. A more realistic CI user (red triangles) might however experience great difficulty in recognising the first target word. As a result, they may need to have the first sentence presented at very high SNR before they recognise the first target word. This would cause the SRT measurement (dotted line) to overestimate the listener's true SRT. In contrast, for a realistic CI user following the modified protocol (blue diamonds), the replacement of sentences following too many unsuccessful trials reduces the risk of divergence of SRT measurements.

A final change was introduced to address CI users' sensitivity to high sound levels. CI users' SRTs were expected to be between 10 and 30 dB higher than NH listeners'. Fixing the noise level to 70 dB and varying the speech level as in the preliminary experiment could have led to uncomfortably loud speech levels for CI users, in excess of 80 dB. Instead the new protocol followed Culling et al. (2012) in maintaining the overall sound level throughout an experiment at 65 dB (A), a level chosen as a normal speech level. A digital sound-level meter was used to set the level correctly whilst the RMS power of the sum of target and masker was kept constant by the software.

4.3.4. Validation of the new SRT protocol with NH listeners

200 high-predictability SPIN sentences were recorded for this test, as defined by Kalikow et al. (1977), with a male English voice and grouped as 20 lists of 10 sentences. The waveforms of the sentences were normalised in RMS level.

The standard protocol made use of the SPIN sentences in a manner similar to that of section 3.2.3, following a 1-up, 1-down adaptive-threshold method. The success of a trial was determined by correct identification of the key word and from the start of the staircase, sentences were presented only once. So as to isolate the impact of sentence repetition in the modified protocol, the standard protocol also made use of a fixed overall sound level of 65 dB SPL. The modified protocol was as described in Section 4.3.3 above.

4 young normally hearing adults were recruited from the Cardiff University undergraduate population, aged between 19 and 22. Each participant was tested in the collocated configuration with the 20 lists. 2 participants were tested with 10 lists in the standard protocol followed by 10 lists in the modified protocol. The order was reversed for the other 2 participants. This resulted in obtaining 40 SRTs per method.

The output of the standard protocol for each run was, as before, the average of the last 8 computed SNRs in the adaptive phase, i.e. the average of the 4th to the 11th computed SNRs. The modified protocol, although run for each participant with the full 10-sentence lists, could be compared to the standard protocol by varying the number of SNRs used (hence the number of sentences used) in the SRT computation. For instance, when computing an SRT using only 9 sentences, the SNRs calculated after presentation of the 10th sentence were neglected. Only SNRs differing from the preceding presentation's SNR were taken into account. The hypothetical outcome shown in Figure 4.2 makes use of 9 sentences and only the SNRs of the filled-symbol presentations were retained in the SRT computation.

Protocol (number of sentences used)	**Mean SRT (dB)**	**SRT standard deviation (dB)**	**Mean number of presentations in the adaptive phase**
Standard (10)	-12.3	1.9	10
Modified (8)	-13.2	2.0	11.5
Modified (9)	-13.3	1.9	13.2
Modified (10)	-13.3	1.7	14.9

Table 4.1: Variability compared between standard and modified protocols

To compute a standard deviation of SRTs that reflected only the replicability of the measurements, inter-participant variation was factored out. To achieve this, the participants' individual SRTs were normalised, so as to equalise each participant's mean SRT to the overall mean SRT. Table 4 compares mean SRTs, the normalised-SRT standard deviation and the mean number of presentations in both protocols. It can be seen that the standard deviations are approximately the same as for the standard protocol when making use of 9 sentences per list in the modified protocol. Perhaps due to the sentence repetitions in the modified protocol, the modified method gave rise to a slight downward shift in SRT. As a result, the mean SRTs differed by 1 dB between methods. Since our primary interest is to measure SRMs, which are relative SRTs, a slight offset in absolute SRTs was not considered an issue. As the modified protocol was thought better adapted to CI users than the standard, it was assumed that running the same test with CI users would lead to a larger variance with the standard protocol. Thus it was decided to use the new protocol with lists that were 9-sentences long.

4.3.5. Modifications to the free-head task protocol

In the preliminary study, listeners were simply required to flag when they had lost track of the clips presented. This provided a subjective measure of SRT. A high level of variability was found in this measure. In an attempt to reduce variability, it was felt that a more precise measurement of when listeners had actually lost track of the speech was required. To that end, listeners were instructed in the refined protocol to recall, immediately after the clip playing had stopped, the last 3-5 words that they felt they had correctly understood in sequence. The clip time and corresponding SNR would subsequently be identified in the clip's transcript to work out a (somewhat less) subjective SRT. As enough material was generated to create longer video clips than those used in Chapter 3, the SNR diminution rate was decreased from 7.5 to 6 dB/min. It was hoped this would help further reduce the variability in subjective SRTs.

The listeners were not told where the target speech would come from. Indeed the presence of the video monitor was sufficient for listeners to spontaneously face it at the start of each trial.

Finally, the free-head task was also run in the collocated, T_0M_0 configuration. Adding this configuration enabled us to obtain a reference subjective SRT for each listener. Subtracting subjective SRTs obtained in spatially separated configurations from the reference SRT would lead to subjective measures of SRM that we could then directly compare with BRIR predictions.

4.3.6. Stimuli selection and preparation

A set of 320 high predictability SPIN sentences were audio-visually recorded with an English male speaker for audio and AV SRT measurements. Additional high predictability sentences were generated, following the rules established by Kalikow et al. (Kalikow et al. 1977). The video recordings were such that the speaker's face covered two thirds of the 17-inch video screen height, delivering a near life-size face. The speaker faced the camera at all times, with his face well lit for lip-reading purposes. The AV files were batch-processed with ffmpeg (ffmpeg.org) to split them into audio (.wav) and video (.mp4) components for separate audio treatment and in-line alteration of sound level during the SRT adaptive tracks. All audio files were equalised for RMS power computed over the 3-4 second recordings. Masker and target audio streams were manipulated to be distributed according to the spatial configuration and desired SNR. When presented in AV mode, the target speech audio and video streams were merged and synchronised by VLC for presentation on the video monitor.

For the free-head listening tasks, the reading of sections of the Project Gutenberg EBook of The Wonderful Wizard of Oz (L. Frank Baum) was audio-visually recorded as per the SPIN sentences. A set of six 6-minute-long video clips was generated. The source material was chosen for its relatively simple vocabulary, high recurrence of words and high predictability. The audio and video components were split as per the SPIN sentences and each 3-4 second segment of the audio stream was normalised for RMS power. In order to ensure that long gaps in speech were not included in the RMS calculation, a threshold level was applied to a 100-ms sliding measurement window with segments falling below this threshold rejected from the calculation. The resulting clips are referred to below as the WizOz clips.

4.3.7. Task sequencing and condition rotation

A first set of free-head listening runs was performed undirected, a second with instruction to explore the benefits of head orientation. Each set consisted of two blocks of 3 presentations (one presentation per spatial configuration T_0M_0, T_0M_{180} and T_0M_{90}). One block was in audio-only and the other in AV mode. Every other participant started with the audio-only block, the remainder with the AV block. Material order remained fixed for all participants and configuration was rotated every other pair of participants.

The SRT measurements followed the modified SRT protocol detailed in Section 4.1. The 5 selected configurations were H_0M_0, H_0M_{180}, $H_{30}M_{180}$, H_0M_{90} and $H_{30}M_{90}$. SRTs in these five configurations were measured in blocks. Alternate participants began with the audio-only or AV block. The order of the sentence lists remained constant for all participants. Within each block, spatial configurations were rotated every other pair of participants. Two repeat runs were performed and SRTs subsequently averaged between repeats.

A shortfall of the preliminary experiment was that conditions were not balanced symmetrically about the median plane. As a result we could not establish or compensate for a plausible bias towards turning one's head one way rather than the other. To correct that oversight, NH listeners performed the T_0M_{90} free-head and SRT task with the masker presented either from the right or the left loudspeaker, allocated alternatively for each participant.

4.4. Results

All model predictions used below were derived from BRIRs acquired in the Cardiff University test room. We will refer to them as 'model predictions' (see Figure 4.2).

4.4.1. Free-head-listening experimental outcomes

Since runs were performed in the collocated, T_0M_0 configuration as well as in the separated configurations, head orientation tracks could be transformed so that the end point of a track represents a subjective SRM measurement. This was achieved by subtracting all SNRs in the separated configuration from the subjective SRT achieved in the corresponding collocated condition. As listeners made a subjective judgement of when they could no longer understand the speech, the last point of each track then represented a subjective measure of SRM achieved at the final head orientation. With SRMs referenced to the collocated subjective SRT of the same presentation modality, the

AV subjective SRM will not incorporate the LRB. However, plotting head tracks in terms of SRM allows them to be presented alongside model predictions.

4.4.1.1. Spontaneous head orientations and subjective SRM

Whilst being kept naïve about our interest in head orientation, 12 young adult listeners performed the first set of free-head orientation runs. Figure 4.4 shows example head-orientation tracks in the symmetrical T_0M_{180} and Figure 4.5 in the asymmetrical T_0M_{90} configurations. Since half of the listeners were presented with the masker to the right in the T_0M_{90} configuration, their head orientations were reflected about the 0º point, so that all tracks could be plotted on the same graph. The SRMs reached correspond well with model predictions for each configuration (pink bands). In 45% of the trials overall, young NH adults were found to spontaneously make use of head orientation.

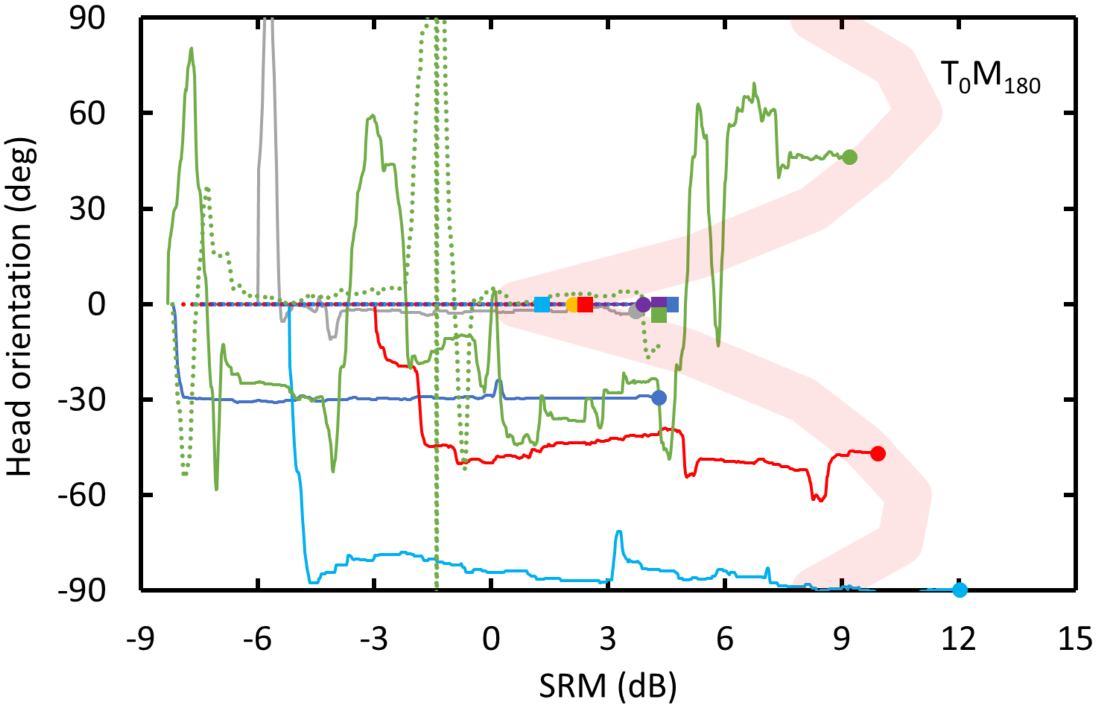


Figure 4.4: T_0M_{180} audio (dotted lines / circles) and AV (solid lines / diamonds) pre-instruction head tracks, subjective SRMs and predictions (pink bands)

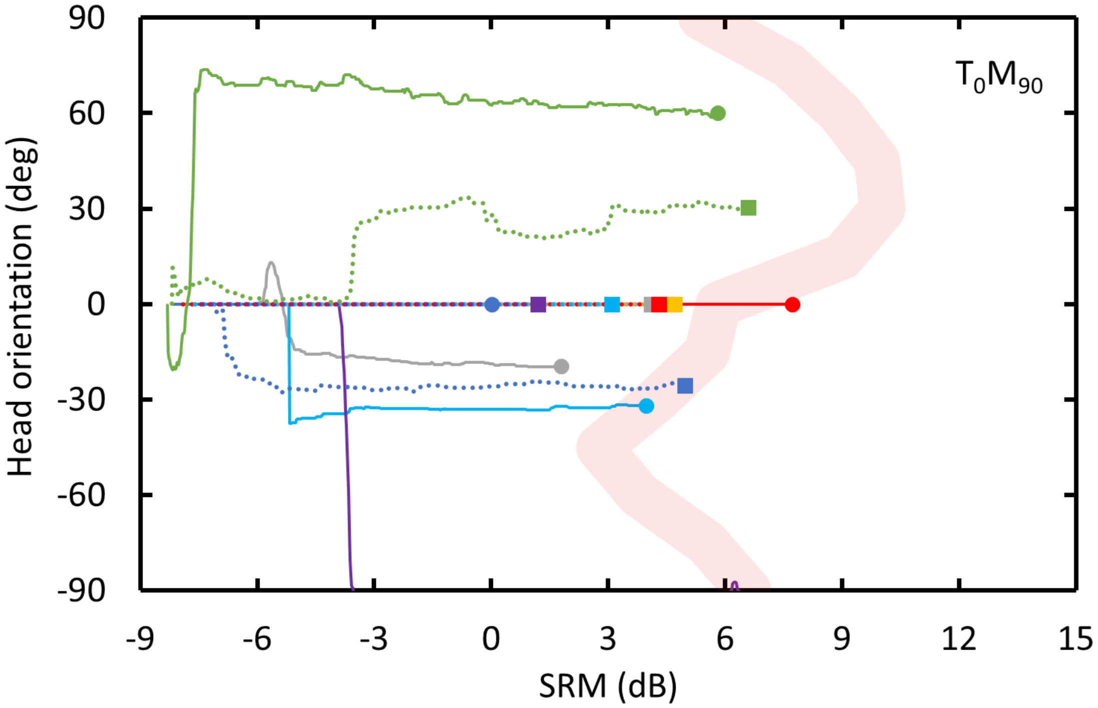


Figure 4.5: T_0M_{90} audio (dotted lines / circles) and AV (solid lines / diamonds) pre-instruction head tracks, subjective SRMs and predictions (pink bands)

For AV presentations, there was a reduction in the use of head turns, indicating that seeing the speaker inhibited the tendency to orient one's head away from them. This was confirmed by comparing the unsigned, wrapped head angle averaged over each track. This measure of temporally averaged head displacement showed a significant main effect of presentation mode ($F(1,11) = 22.28$, $p < 0.002$), but no effect of masker separation ($F(1,11) = 0.87$, $p > 0.4$), nor any interaction between masker separation and presentation modality ($F(1,11) = 0.14$, $p > 0.8$).

SRM was positively correlated with the listeners' final unsigned head orientation in the T_0M_{180} configuration ($r = 0.67$, $t(22) = 4.19$, $p < 0{,}001$), indicating that this benefit is related to orienting away from the speech source. However, no significant correlation was found between SRM and listeners' final head orientation in the T_0M_{90} configuration ($r = -0.36$, $t(21) = -1.76$, $p = 0{,}094$). This poor correlation was primarily due to a large proportion of participants not moving their heads in the AV modality.

Comparing subjective SRM outcomes, despite AV presentation causing a reduction in head movements, it had no significant effect on SRM ($F(1,11) = 1.69$, $p > 0.2$). The mean subjective SRM of 3.8 dB reached in the T_0M_{180} configuration was only 0.8 dB larger than in the T_0M_{90} configuration, and this difference was non-significant ($F(1,11) =$

4.608, p = 0.055). No interaction between configuration and presentation mode was found.

4.4.1.2. Effect of instruction

Once listeners were explicitly informed that orienting their head may be helpful with the task, a second set of runs was performed with the same listeners. Figures 4.6 and 4.7 plot the resulting new tracks.

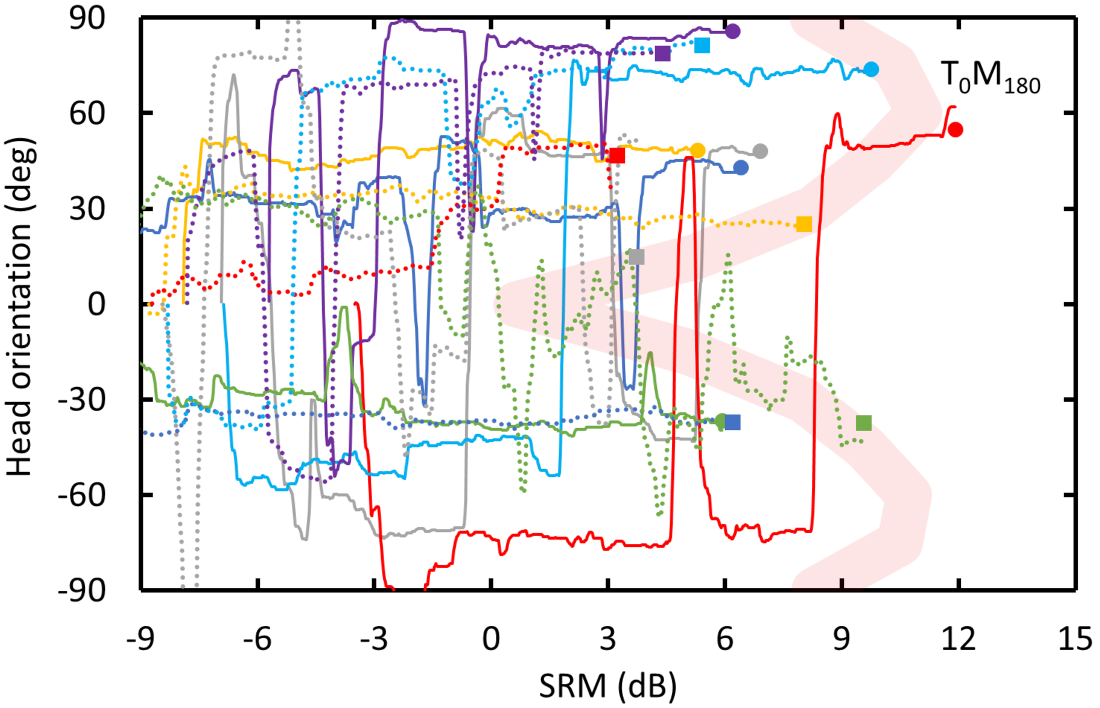


Figure 4.6: T_0M_{180} audio (dotted lines / circles) and AV (solid lines / diamonds) post-instruction head tracks, subjective SRMs and predictions (pink bands)

Instruction unsurprisingly gave rise to an increase in head movements when measured as the unsigned, wrapped head angle averaged over each track ($F(1,11) = 73.32$, $p < 0.001$). Subjective SRMs significantly rose as a result of increased head movements by an average of 1.6 dB compared to the undirected runs ($F(1,11) = 7.80$, $p < 0.02$), to 6.4 and 4.6 dB at T_0M_{180} and T_0M_{90} respectively. An immediate benefit to intelligibility in noise was thus observed from this simple instruction to explore the benefit of head orientation. Overall, AV presentation continued to significantly inhibit head movements ($F(1,11) = 35.88$, $p < 0.001$). No significant bias towards turning one way rather than the other was found in the T_0M_{180} configuration, suggesting that handedness did not impact the direction of head turns.

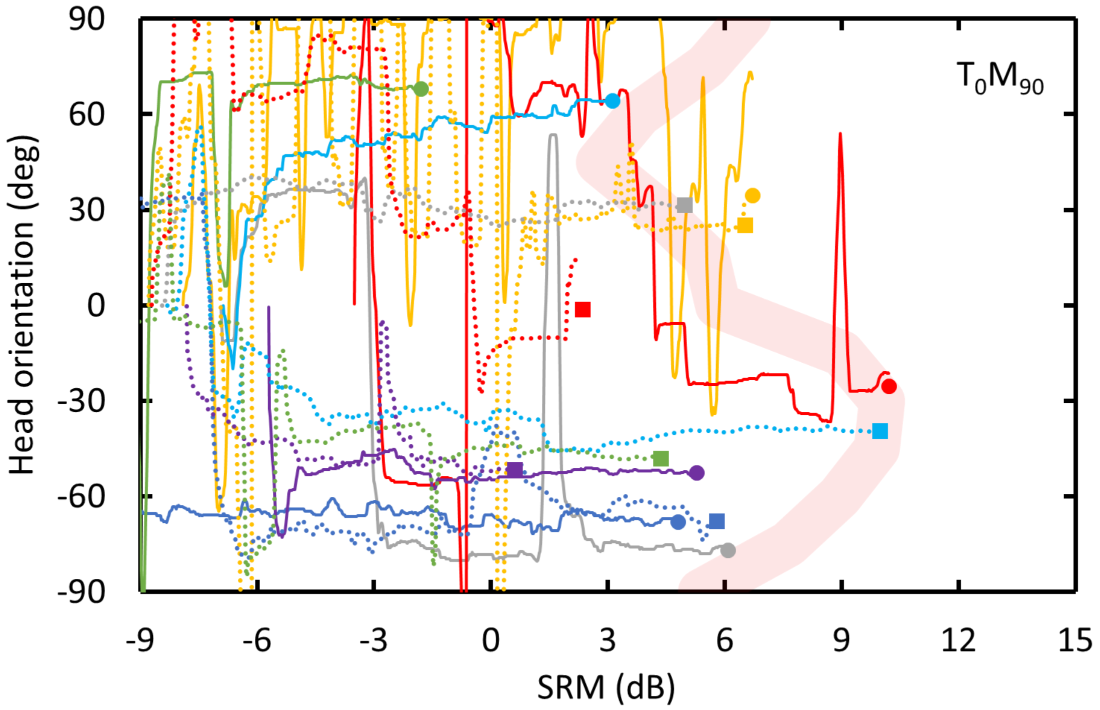


Figure 4.7: T_0M_{90} audio (dotted lines / circles) and AV (solid lines / diamonds) post-instruction head tracks, subjective SRMs and predictions (pink bands)

Post-instruction alone, AV presentation significantly reduced the amount of head movements ($F(1,11) = 15.3, p < 0.002$), but it had no significant effect on subjective SRM ($F(1,11) = 0.8, p > 0.35$). Spatial configuration however had a significant effect on subjective SRM ($F(1,11) = 5.77, p = 0.035$) post-instruction. This is understandable, since any head turn away from the speech released HOB at T_0M_{180} whilst only turns in the correct direction did at T_0M_{90}. The configuration effect was not significant pre-instruction because listeners had just about compensated for a lower T_0M_{180} SF-SRM with always positive HOBs. Post instruction, larger head turns released more SRM at T_0M_{180} than at T_0M_{190}, not only because some listeners persisted in turning the wrong way, but also because some overshot the optimum 30º orientation at T_0M_{190}.

4.4.1.3. Subjective SRM vs. model predictions at final head orientations

Figure 4.8 combines pre- and post-instruction objective SRM data points plotted against the predicted SRM for final head orientations in the T_0M_{180} and T_0M_{90} configurations. Where extreme head orientations were adopted, participants may not always have made use of lip-reading in the AV modality. Hence, only the audio data is considered here.

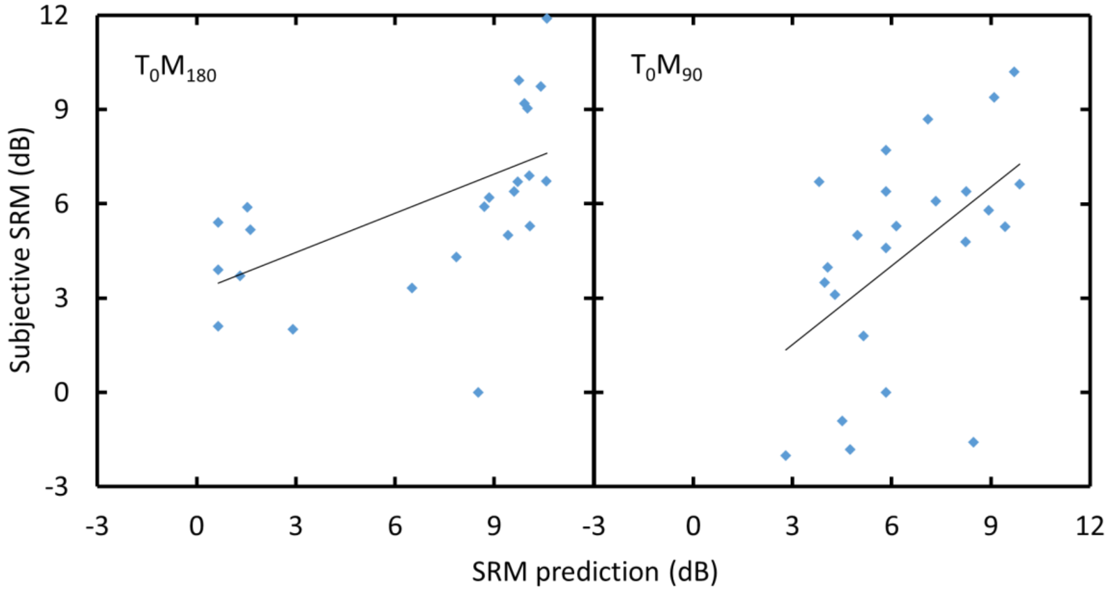


Figure 4.8: Correlation between subjective SRMs and SRMs predicted for final head orientations in the T_0M_{180} and T_0M_{90} configurations

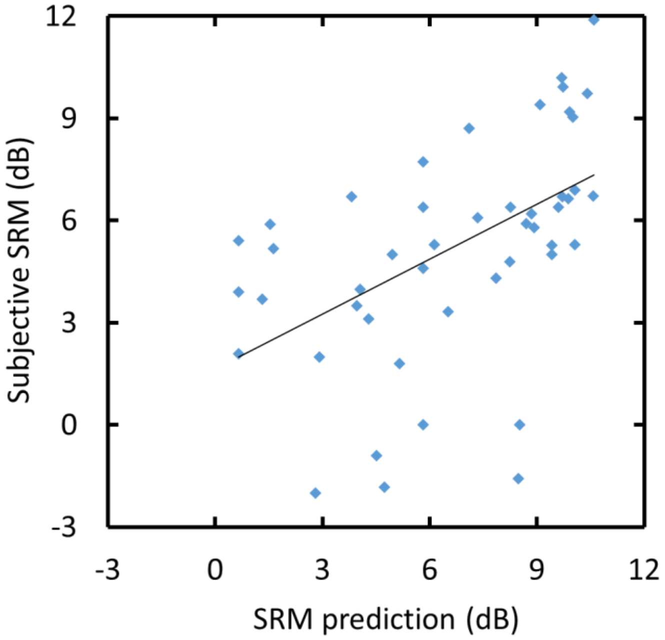


Figure 4.9: Correlation between subjective SRMs and SRMs predicted for final head orientations across configurations

Significant correlations were found between the subjective SRM data and predictions at T_0M_{180} (r = 0.53, t(22) = 2.92, p < 0.01, slope 0.42 with 0.14 SE) and at T_0M_{90} configuration (r = 0.51, t(22) = 2.76, p = 0.011, slope 0.84 with 0.30 SE). This confirms that the more listeners exploited head orientation the more they could understand of the clip.

Combining the two spatial configurations led to an overall significant correlation between subjective SRMs and predictions (r = 0.49, t(46) = 3.86, p < 0.001) with a 0.54 slope (0.14 SE), as illustrated in Figure 4.9. This is somewhat less than the 0.86 slope found in Chapter 3's spatial configurations but the large variability of the data (quite apparent in the scatter plot) we believe explains the discrepancy.

4.4.2. Objective-SRT experimental outcomes

10 young normally hearing listeners performed the (fixed-head) SRT task. The speech-facing SRM (SF-SRM) and additional 30º HOB are plotted in Figure 4.10 for audio-alone and for AV presentations against model predictions for the two spatially separated configurations.

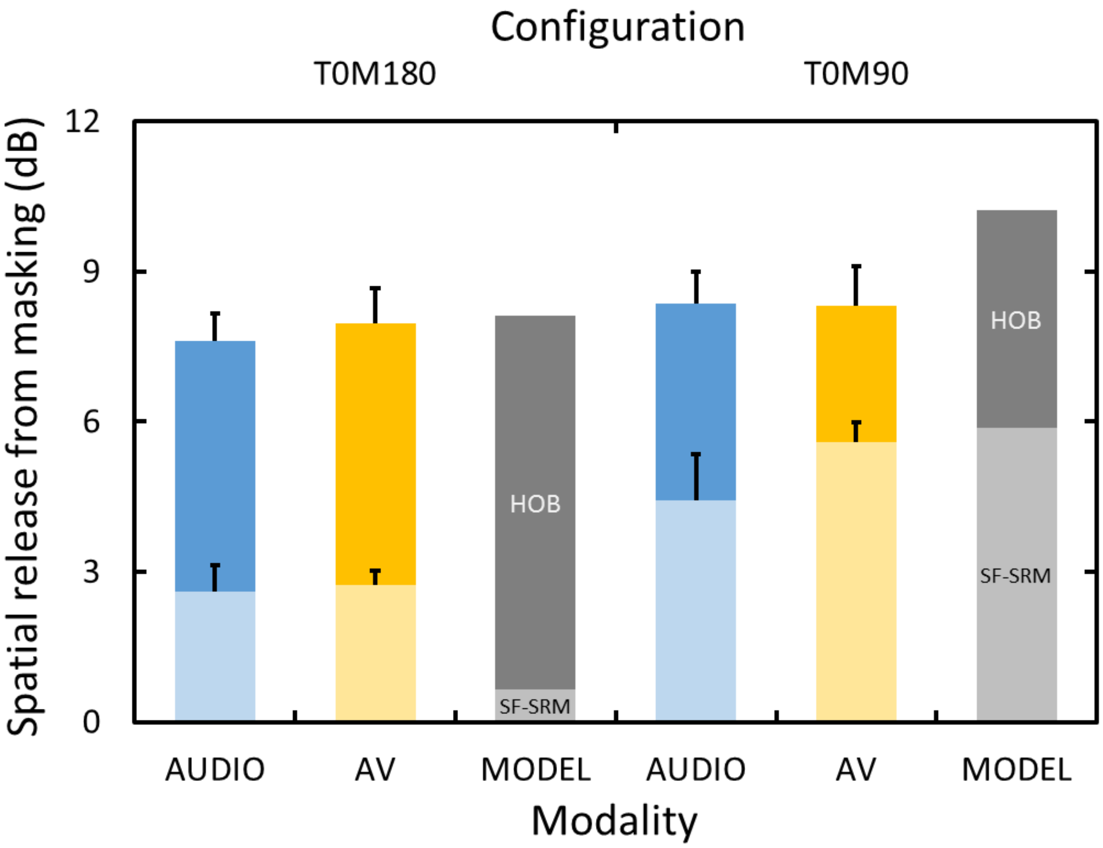


Figure 4.10: Cumulative effect of SF-SRM (pale lower bars) and 30º HOB (dark upper bars) in audio and AV modalities against model predictions. Error bars are standard error of means

SF-SRMs were computed by subtracting the H_0M_{180} and H_0M_{90} SRTs from the H_0M_0 SRTs whilst HOBs were computed by subtracting the H_0M_{180} from the $H_{30}M_{180}$ SRTs and the H_0M_{90} from the $H_{30}M_{90}$ SRTs. This calculation was performed within each modality such that the LRB was cancelled out in the AV results. Figure 4.10 makes use of stacked columns since the SF-SRM and HOBs are cumulative. The variability of outcomes is represented here, as in Figure 4.11, as standard error of means (SE) error bars.

4.4.2.1. Speech-facing SRM

SF-SRM data in H_0M_{180} and H_0M_{90} configurations are displayed as the pale lower bars of Figure 4.10 for audio alone and for AV presentation modalities. The 2.6 dB H_0M_{180} audio SRM was large compared to the BRIR-based prediction of 0.66. However, much of that discrepancy can be explained by participants deviating from facing the speech during measurements. In Chapter 3 we saw how a deviation as small as 5º in either direction could account for at least 1 dB HOB at T_0M_{180}, where the HOB slope is maximum around facing the speech. The T_0M_{180} SF-SRM in AV modality was almost identical to the audio case.

For T_0M_{90}, audio and AV SF-SRMs were 4.4 and 5.6 dB (0.93 and 0.39 dB SE) respectively. This compares with a model prediction of 5.8 dB. These T_0M_{90} SF-SRMs are somewhat lower than the 7 dB found for young normally hearing listeners in Culling et al. (2012). However, Culling et al.'s T_0M_{90} SF-SRM was 1 dB larger than our model prediction because their test room was a little less reverberant than ours.

An ANOVA for speech-facing SRT confirms that masker separation had a significant effect on SF-SRT ($F(2,18) = 50.18$, $p < 0.001$), with pairwise comparisons between H_0M_0, H_0M_{180} and H_0M_{90} all showing significant differences in SRTs ($p < 0.002$) and therefore significant SF-SRM.

4.4.2.2. Additional 30º HOB

The HOB obtained in T_0M_{180} and T_0M_{90} configurations can be seen in Figure 4.10 (dark upper bars), for audio-alone and for AV modalities. At T_0M_{180}, HOBs of 5.0 and 5.2 dB were obtained in audio and AV, respectively (with 0.6 and 0.7 dB SE). This compares to a model prediction of 7.5 dB. The references for these HOBs are the SRTs obtained at H_0M_{180} which we assumed were improved (i.e. SRTs lowered) by listeners' unintentional deviation from facing the speech as they focussed on the listening task. At T_0M_{90}, 3.9 and 2.7 dB were gained from head turns in audio and AV respectively (with 0.7 and 0.8 dB SE), compared to 4.4 dB predicted.

An ANOVA for SRM within each presentation modality confirmed that head orientation had a significant, beneficial effect ($F(1,9) = 108.76$, $p < 0.001$).

The cumulated SRMs (from combined masker separation and head orientation) in the T_0M_{180} configuration were 7.6 and 8.0 dB in audio and AV respectively, very close to the 8.1 dB model prediction. This good data match with predictions strengthens the assumption that listeners did indeed deviate from facing the speech at H_0M_{180}.

4.4.2.3. Lip-reading benefit

NH listeners' LRB was computed by subtracting the AV mean SRT from the audio-only mean SRT. Figure 4.11 displays the benefits measured in each spatial configuration. The LRB ranged across participants and configurations from -0.7 to +5.5.4, averaging 3 dB. The cross-participants means ranged from 2.6 to 3.8 dB across configurations.

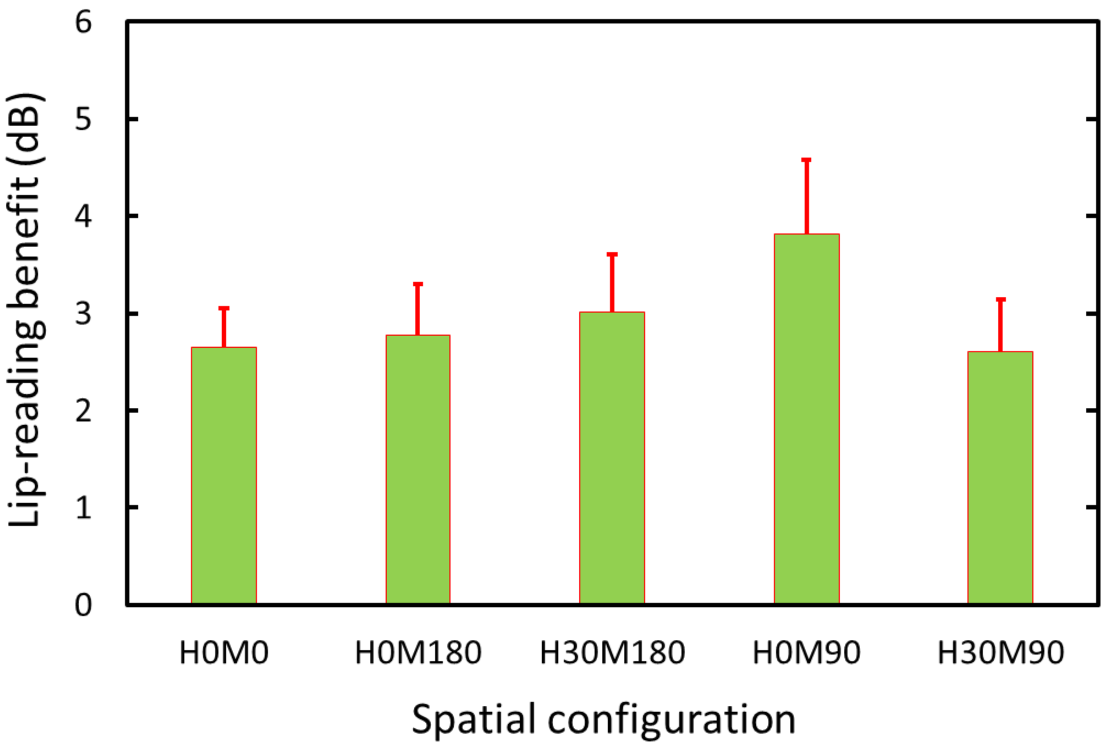


Figure 4.11: Young NH adults' LRB. Error bars are standard error of means

An ANOVA for SRTs in the two presentation modalities and across the 5 overall spatial configurations confirmed a significant benefit of visual cues ($F(1,9) = 68.58$, $p < 0.001$). There was no interaction between modality and configuration ($F(4,36) = 0.83$, $p > 0.5$), indicating that configuration did not have a significant effect on LRB. Most relevant to our study, a 30º head turn had no detrimental effect on LRB ($F(1,9) = 0.19$, $p > 0.19$).

4.5. Discussion

Objective SRM of young normally hearing listeners was measured as a function of both masker separation and head orientation. With the target directly ahead and a masker placed behind (T_0M_{180}) or to the side (T_0M_{90}) of the listener, SRTs compared to those measured in the collocated (T_0M_0) configuration led to audio-only and AV SRMs of 7.6-8.3dB being reached thanks to a modest (30º) head turn away from the speech. This matched model predictions within less than 0.5 dB and 1.9 dB in the T_0M_{180} and T_0M_{90} configurations, respectively. In addition, listeners reached, on average, 3 dB lower SRTs when the speaker's face was visible and this 3 dB LRB was unaffected by a 30º head turn. 30º HOBs of 5.1 and 3.3 dB were measured in the T_0M_{180} and T_0M_{90} configurations respectively, somewhat less than the predicted 7.5 and 4.3 dB, but that discrepancy was accounted for by the listeners experiencing difficulty with accurately facing the speech when SRT measurements required it.

When attending to a diminishing speech level in a fixed noise level and in the same three spatial configurations, listeners in an undirected paradigm were found to make use of head orientation to improve their speech intelligibility in 45% of trials. That was less than the 60% found in Chapter 3, but this can be explained by a clear reduction in spontaneous head movements when the speaker's face was made visible. As was found in Chapter 3, those who spontaneously moved their heads used a range of strategies that could not be clearly categorised. The objective SRT data shows that with a 30º head turn, 7.8 and 8.3 dB SRMs were available to listeners in the T_0M_{180} and T_0M_{90} configurations respectively. However, they achieved subjective SRMs of only 3.8 and 3 dB, respectively, in the free-head task. This shows how relatively poor listeners were at spontaneously making use of an effective head-orientation strategy. No significant bias in head orientation (to the left or to the right of the speech direction) was found in the T_0M_{180} configuration. However, in the T_0M_{90} configuration many listeners spontaneously turned their heads in the direction opposite to the noise direction, as though to get away from the masker, but the wrong way in order to obtain a HOB. This explains why they gained on average only 0.5 dB HOB over those who did not turn their heads, very small compared to an objective 30º HOB of 3.3 dB. After instruction to make use of head orientation, although head movements significantly increased, listeners still only reaped an extra 1.6 dB HOB out of the > 4.2 dB available to them. This clearly shows that young normally hearing listeners remained poor at making use of effective head-orientation strategy, as was found in Chapter 3, and would benefit from advice about *how* to move their heads.

On average, audio-only subjective SRMs in the T_0M_{180} configuration were marginally lower (at 6.1 dB) than those predicted from the final head orientations (7.0 dB). In the T_0M_{90} configuration, subjective SRMs averaged 4.4 dB compared to 6.4 dB predicted. Similar effects were seen in the objective measures of SRM (though measured only at two head orientations) when compared to predictions.

The lip-reading benefit was measured objectively as 3 dB. Our intention was not to quantify the benefit of visual cues as much as it was to verify that, whilst a modest 30º head orientation would significantly improve speech intelligibility in noise, it would not significantly affect the listeners' lip-reading ability. Previous studies, however, examined the contribution of vision to speech perception in noise (e.g. MacLeod & Summerfield 1987; Macleod & Summerfield 1990; Summerfield 1992). MacLeod and Summerfield developed an adaptive method of measuring audio-alone and AV SRTs based, as ours, on the efficient technique developed by Plomp & Mimpen (1979). Making use of BKB sentences, they typically measured LRB between 6 and 15 dB among young, normally hearing subjects and from 3 to 22 dB using sentences (MacLeod & Summerfield 1987). The difference between sentence corpora used then and here, coupled with our speaker being reported by CI users as relatively hard to lip-read, most likely explain our measure of LRB being as low as 3dB. A hard-to-lip-read speaker might enhance the difference between normally hearing adults, naturally poor lip-readers, and CI users, known to rely much more heavily on lip-reading (Mitchell & Maslin 2007). The reverse might be true, though. The superior LRB of CI users will be further explored in Chapter 5.

4.6. Conclusion

The pilot experiment presented in this chapter further validated the Jelfs et al. (2011) model of SRM for predictions of head-orientation benefits in audio-only and AV modalities. In moderately reverberant conditions, objective measures of SRT showed that a 30º HOB could reach up to 5.1 dB for young normally hearing listeners, without compromising the 3 dB benefit they obtain from lip reading. In a free-head paradigm listeners were relatively poor at reaping the HOB available to them. In 55% 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 or erratically. They generally performed poorly, reaping on average only 3.4 of the 8 dB available to them at a 30º head orientation (the additional 2.7 dB SRM available with a

60 ° orientation at T_0M_{180} would have cost them the bulk of their 3 dB LRB). Some listeners went as far as choosing head orientations that made speech intelligibility worse even though they might have passed through beneficial head orientations and/or localised the masker position. Repeating the task after instruction, listeners generally improved, although there was a lot of room for further improvement. This indicated that training on how to optimise one's head orientation strategy could be beneficial, but that young normally hearing listeners were not experts at this task and could benefit a lot further from more specific advice on how to move their heads.

5. AUDIO & AUDIO-VISUAL EXPERIMENTS WITH ADULT CI USERS & AGE-MATCHED, NORMALLY HEARING LISTENERS

The objective SRT experiment of Chapter 4 demonstrated how young, normally hearing listeners can reap the speech-in-noise intelligibility HOB (4.2 of 5.9 dB predicted) of a modest 30º head turn away from the speech when the masker is placed behind or to the side of the listener. Testing in both audio-only and AV modalities further showed that young NH listeners also benefit from lip-reading and that such a benefit (3 dB) is unaffected by a 30º head turn. When tasked with following a continuous monologue of diminishing SNR for as long as they could understand it, and in the same spatial configurations as in the SRT tasks, young NH listeners spontaneously turned their heads only 45% of the time and when they did turn, they seldom made effective use of head orientation. On average across conditions, listeners spontaneously reaped only 3.9 dB (subjective) SRM of the 8 dB (objective) SRM available to them through a 30° head turn, 3.3 dB of which was objectively measured as SF-SRM. Seeing the speaker's face significantly reduced spontaneous head turns when compared to audio-only presentation. After being instructed to explore the benefits of head orientation, listeners improved, but remained poor at the task since they subjectively reaped only an extra 1.6 dB SRM on average. Some individuals performed even more poorly after instruction, moving to negative SRM head orientations, even though they had already been exposed to the material during the undirected task.

The experiments in Chapter 4 allowed us both to validate paradigms designed for CI users and to obtain a young NH baseline. We were then ready to explore whether adult CI users can gain substantial HOBs. We also needed to confirm that as for NH listeners, a 30º head turn did not adversely affect their LRB. We also examined whether CI users spontaneously turn their heads in the free-head task and how much they benefitted from being instructed to explore the effects of head orientation.

5.1. Spatial configurations, predictions and hypotheses

The spatial configurations and head orientations used in Chapter 4 were used again, but here to test three different groups of listeners: bilateral CI users, unilateral CI users and NH listeners who were age-matched to the CI users.

Bilateral CI users who perceive speech equally well through either CI would always be expected to benefit from head-shadow. In the T_0M_{90} configuration, they would benefit regardless of the side to which the noise was presented. In the T_0M_{180} configuration they would benefit regardless of the side to which they turned their heads. However, as most adult bilateral CI users in the UK perceive speech better through one of their CIs, our participants were quizzed prior to testing as to which CI they could best hear speech with. That CI would be considered their 'better ear' and in the $H_{30}M_{180}$ and $H_{30}M_{90}$ SRT runs, only the 30° head turn favouring their better ear would be tested for HOB. As illustrated in Figure 5.1, such a 30° head orientation is predicted to bring their better ear closer to the target speech and to keep it or bring it in the shadow of the head with respect to the masker. Both changes contribute to increasing the SNR at the better ear. In the T_0M_{90} configuration, an additional SNR increase results from moving the better ear away from the masker bright spot (see Sections 1.2.2 and 2.4.2). In Figure 2.4, the masker bright spot in the target-in-front situation is represented by the kink in the SRM curve (or SRM valley) at or around 90° masker position. Escape from the bright spot (with a 90° masker azimuth) with head rotation away from facing the speech is clearly shown in Figure 2.5 for anechoic conditions. There, and for a left better ear, one can move from the SRM valley, situated around the speech-facing orientation, to the largest peak 30° to the right or to a smaller peak 10-20° to the left. The fact that moving the better ear away from the speech is predicted to provide a small but measureable HOB may be surprising. It is in fact symptomatic of the bright spot being strong enough to locally reverse the SRM trend that one would obtain, were the bright spot absent. The bright spot effect is eroded by the little reverberation we have in our sound-deadened room. This can be seen in Figure 5.3 below. In our testing room, the kink in the SRM curve is still a pronounced valley when the noise is moved 10 or 20° beyond 90°, as was illustrated in both data and predictions in Figure 3.13, but for the selected conditions in this and the previous chapter, no SRM valley is expected as such.

In the H_0M_0 configuration, where speech and noise are collocated and in front, the input to both ears is the same. In this case, NH listeners can benefit from a summation effect. Bilateral CI users are also known to benefit from a summation effect (Schleich et al. 2004). By additionally measuring the bilateral CI users' H_0M_0 SRTs with each CI disabled in turn, we set out to measure summation. It is computed by subtracting the SRT obtained with both implants enabled from the SRT obtained with their better ear. This measurement also allowed us to confirm that the perceived better ear for speech perception in noise of a given participant was indeed their better ear.

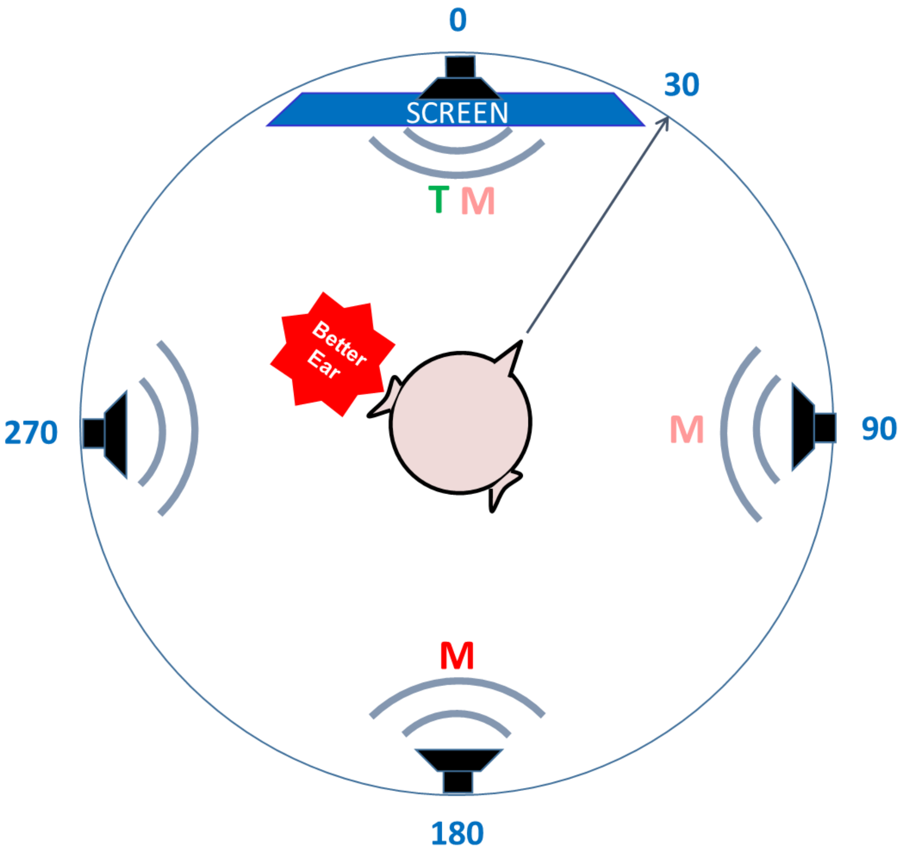


Figure 5.1: Schematic of the SRT spatial configurations. Codes incorporate masker separation and head orientation (highlighted is the $H_{30}M_{180}$ configuration)

In spatially separated configurations where the input to both ears differs, the brain selects, in each frequency band, the signal receiving the best SNR. This leads to the 'better ear' or 'head shadow' effect, where the signal to the poorer ear appears to be discarded. A different input in both ears also leads to binaural unmasking. The comparison by a normally hearing listener's brain of the fine structure of the sounds arriving at each ear enables rejection of some of the noise and better intelligibility of speech. This benefit is sometimes measured by subtracting SRTs with the better ear from those from using both ears, but this time in a spatially separated situation. Bilateral CI users have little to no access to the fine structure of sounds because of the sound processing strategies commonly employed. Yet, bilateral CI users have also been shown to benefit from such an addition of the worse ear. This effect is known in the CI literature as "squelch" (Schleich et al. 2004). Since CI users have no access to fine structure information, the effect must be presumed to originate from different mechanisms in their case. In order to fully understand the HOB of bilateral CI users when it is compared to model predictions that only take into account the BE contribution to SRM, it is necessary to additionally measure the participants' squelch. We set out to do so in two spatially separated configurations: H_0M_{90}, the standard configuration for squelch measurement in the literature, as well as $H_{30}M_{180}$.

The HOBs of bilateral and unilateral CI users were measured in the T_0M_{90} configuration that favours their better ear (or their sole CI in the case of unilateral users). In this case, the HOB can be compared with the NH controls. Unilateral CI users were predicted to have markedly elevated SRTs in the T_0M_{270} configuration, where the masker faces their CI, but to still have a HOB. We also tested the HOB of unilateral CI users in that situation in order to test these predictions. The additional configuration is illustrated in Figure 5.2. BRIR-model predictions in all three configurations and for each listener group are plotted in Figure 5.3.

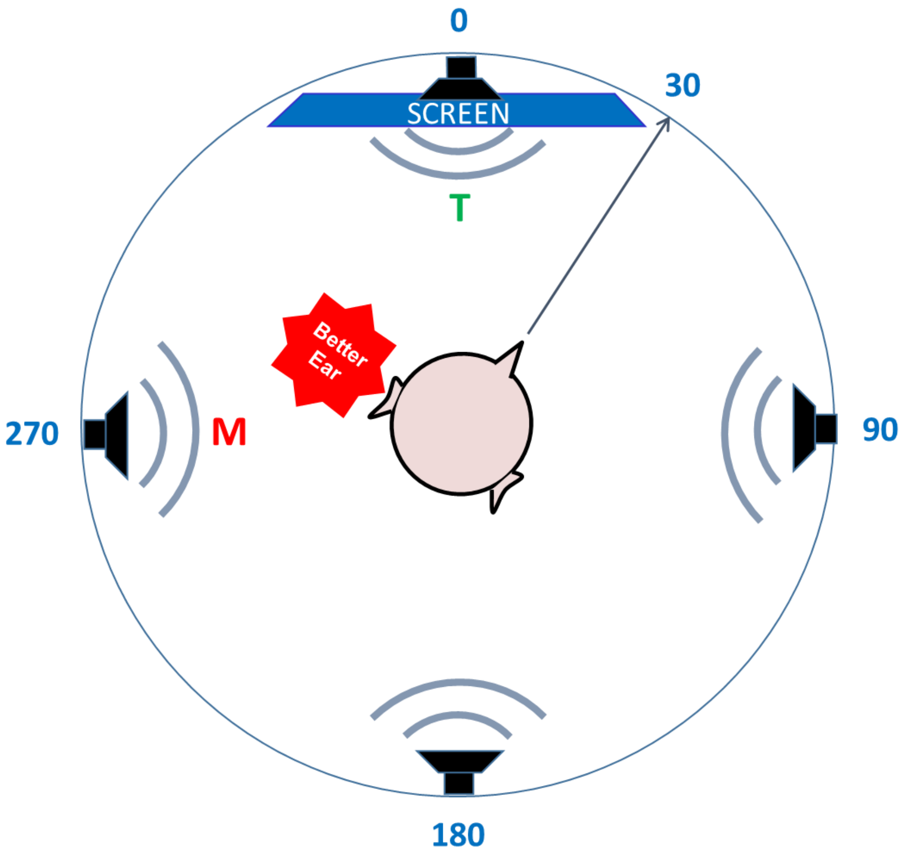


Figure 5.2: Schematic of the additional T_0M_{270} SRT spatial configuration for unilateral CI users. The highlighted 30° head turn ($H_{30}M_{270}$ configuration) is predicted to provide unilateral users with significant HOB

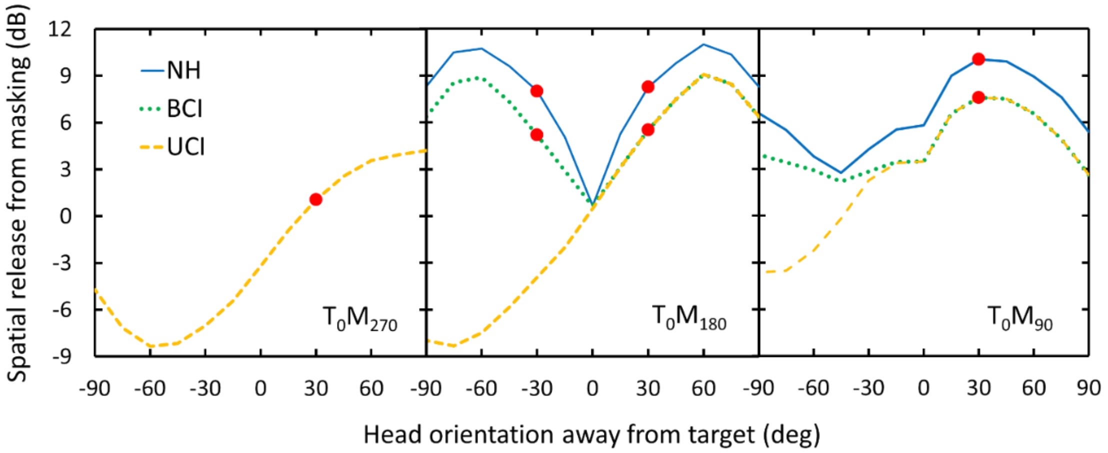


Figure 5.3: BRIR-model predictions in the T_0M_{270}, T_0M_{180} and T_0M_{90} configurations for NH listeners, bilateral (BCI) and unilateral (UCI) CI users. Red dots highlight the predicted SRM at a 30º head orientation

Following the experience of testing NH listeners, described in Chapter 4, we expected measures of SRM to systematically deviate from model predictions. In Chapter 4, listeners appeared to have difficulty exactly facing the speech in the H_0M_{180} configuration. Consequently, it was anticipated that the measured T_0M_{180} SF-SRMs and HOBs would be above and below predictions respectively. Should the bilateral CI users' squelch turn out to be significant, this may also boost the measured SF-SRMs and reduce the HOBs, when compared to model predictions. The measured LRB was expected to be larger for CI users than for NH listeners, as found elsewhere (see Section 1.3.2). As was found for young NH listeners, it was hoped that there would be no significant detriment of a 30º head turn to CI users' LRB.

In the free-head listening experiment, listeners were to be tested in the same configurations as in Chapter 4. For CI users, the T_0M_{90} configuration selected was to favour the listener's better or sole CI. According to predictions, making use of head orientation could provide CI users with a HOB of up to 5 dB and hence could make the difference, in a noisy social setting, between CI users being engaged in a conversation and being isolated from it. We therefore expected CI users to be more motivated to turn their heads than NH listeners and to make more spontaneous use of head orientation in the free-head task. Since unilateral CI users do not benefit from binaural hearing, they more frequently have to make head movements to make an assessment, be it poor, of the localisation of sound sources. For that reason we further hypothesised that unilateral CI users would spontaneously perform the best. As found in Chapter 4, we expected the AV modality to inhibit head movements and probably more so for CI users as they are more reliant on lip-reading. After instruction to make use of head orientation, we expected CI users to perform better than NH listeners, given the poor performance previously found with young NH listeners.

5.2. Materials & Methods

5.2.1. CI participants

8 bilateral CI users and 10 unilateral CI users participated. All CI users were recruited from England and Wales through the National Cochlear Implant User Association (NCIUA) and the Cochlear Implant User Group 2004 (Yahoo! CIUG-2004 group). Data from 9 unilateral users was retained as one user could not understand enough of the speech material, even in silence. The unilateral users were aged between 32 and 74, averaging 58; the bilateral users between 48 and 78, averaging 67. The unilateral users had negligible residual hearing in the non-implanted ear.

CI user	**Age**	**Left CI**				**Right CI**				**Aetiology**
		Year fitted	**Make**	**Processor**	**Implant**	**Year fitted**	**Make**	**Processor**	**Implant**	
B1	78	2013	Cochlear	Nucleus6	CI-500	2013	Cochlear	Nucleus6	CI-500	Unknown
B2	64	1995	MedEl	Tempo+	Pro short-h	2000	MedEl	Tempo+	CIS Pro+	Meniere
B3	48	2005	Cochlear	Nucleus6	N24	2012	Cochlear	Nucleus6	CI24-RE	Genetic
B4	71	2009	AB	Harmony	HiRes90K	2011	AB	Harmony	HiRes90K	Usher
B5	67	2004	Cochlear	Nucleus5	N24	2006	Cochlear	Nucleus5	CI24-RE	Meniere
B6	66	2001	MedEl	Opus2	Combi40+	2005	MedEl	Opus2	Pulsar	Unknown
B7	66	2001	MedEl	Opus2	Combi40+	2001	MedEl	Opus2	Combi40+	Unknown
B8	78	2007	AB	Harmony	HiRes90K	1995	Cochlear	Freedom	N22	Unknown

Table 5.1: Specifics of bilateral CI user participants

Tables 5.1 details the specifics of our bilateral CI participants. All but one bilateral CI user (B1) had had their last implant fitted at least a year prior to testing and had sequential implantation with the second implant fitted between 2 and 12 years after the first. Participant B1 was simultaneously implanted and had the implants switched on 3 months before testing. Table 5.2 details the specifics of our unilateral CI participants. All unilateral CI participants had had their implant fitted at least 3 years before testing.

CI user	**Age**	**Side fitted**	**Year fitted**	**Make**	**Processor**	**Implant**	**Aetiology**
U1	39	Right	2003	AB	Harmony	C2	Sensorineural
U2	60	Left	2010	MedEl	Opus2	Pulsar	Meniere
U3	67	Left	2004	MedEl	Opus2	Combi40+	Unknown
U4	67	Left	2008	AB	Harmony	HiRes90K	Unknown
U5	32	Left	2004	AB	Harmony	HiRes90K	Unknown
U6	74	Left	1996	Cochlear	Nucleus5	N22	Streptomycin
U7	59	Right	2008	Cochlear	Freedom	N24	Unknown
U8	65	Left	1997	Cochlear	Freedom	N22	Unknown
U9	66	Left	2002	Cochlear	Esprit 3G	N24	Viral inf.

Table 5.2: Specifics of unilateral CI user participants

All participants but one (U9) had hardware or software settings such that no microphone directionality was used during testing. Participant U9 used the Esprit 3G processor from Cochlear. Its microphone is directional and favours sounds originating from in front of the CI user and inherently filters out some of the sounds originating from the rear. This participant's data will be treated separately so as to specifically investigate the effect of microphone directionality on SF-SRM and HOB.

5.2.2. Age-matched normally hearing participants

The normally hearing listeners were age-matched to the CI users within ±5 years and recruited from the local Cardiff population. All had normal hearing for their age, which was confirmed via acquisition of audiograms (> -20 dB between 500 Hz and 4 kHz). From the 10 normally hearing listeners, a set of 8 was age-matched to each CI group. Table 5.3 illustrates the age-matching.

NH	**Age**	**Bilateral CI user** (BCI)	**Age**	**Unilateral CI user** (UCI)	**Age**
N1	28			U5	32
N2	48	B3	48	U1	39
N3	56			U7	59
N4	62	B2	64	U2	60
N5	67	B6	66	U8	65
N6	67	B7	66	U3	67
N7	67	B5	67	U4	67
N8	69	B4	71	U6	74
N9	80	B1	78		
N10	82	B8	78		
Mean	**62.6**	**67.8** (age-matched NH)	**67.3** (BCI)	**58.0** (age-matched NH)	**57.9** (UCI)

Table 5.3: NH listeners' age-matching to CI users

5.2.3. Laboratory setup

For ease of access to the study, about half of the CI user testing was performed in Cardiff University, the other half at University College London. This was made possible by the use of mobile laboratory equipment, the detail of which is reported in Section 4.3.2. We endeavoured to ensure sufficient consistency of acoustics between the two sound-deadened rooms. The Cardiff site room (Lab 2.14a, School of Psychology Tower Building) measured 3.2 x 4.3 m whilst the UCL room (Lab F, Chandler House, UCL Speech Hearing & Phonetic Department) measured 2.5 x 3.4 m. The distance between the loudspeakers and the centre of the participants' heads was 1.5 m in Cardiff, 1.2 m in London. In order to obtain a reasonable acoustical match between the two setups, sound-absorbing foam panels were added around the UCL room to cover acoustically reflective electrical trunking, as was done in the Cardiff room. The RT60 of the two sound-deadened rooms was measured from impulse responses using Schroeder's reverse integration technique to be circa 60 ms. BRIRs acquired in both sites with the same B&K head-and-torso manikin were fed into the model to verify that predictions were comparable between sites. An illustration of this for the T_0M_{90} configuration is shown in Figure 5.4. Over all the combinations of CI user type and spatial configurations, the discrepancy between BRIR predictions did not exceed 1.5 dB. The worst mismatch in predictions was for a bilateral CI user at the 0° head orientation in the T_0M_{180} configuration. Elsewhere, the mismatch did not exceed 1.2 dB and, more importantly, did not impact the HOB predictions by more than 0.5 dB. As more participants were tested at the Cardiff site, it was decided that the Cardiff BRIR predictions would be used in all the data analysis.

5.2.4. Experimental protocols and material

All participants were first tested with exactly the same free-head task and SRT protocols as described in Chapter 4. As it was crucial that listeners remain naïve about our interest in head orientation in the undirected free-head task, the free-head runs were performed before the SRT runs. In both experiments the order of the speech material remained fixed whilst conditions were rotated across participants. As per Chapter 4, participants performed blocks of runs containing all the configurations to be tested. Alternate participants began with an audio-only or AV block and configuration was rotated every other pair of participant. The experimental sequencing and condition rotation is illustrated in Figure 5.5.

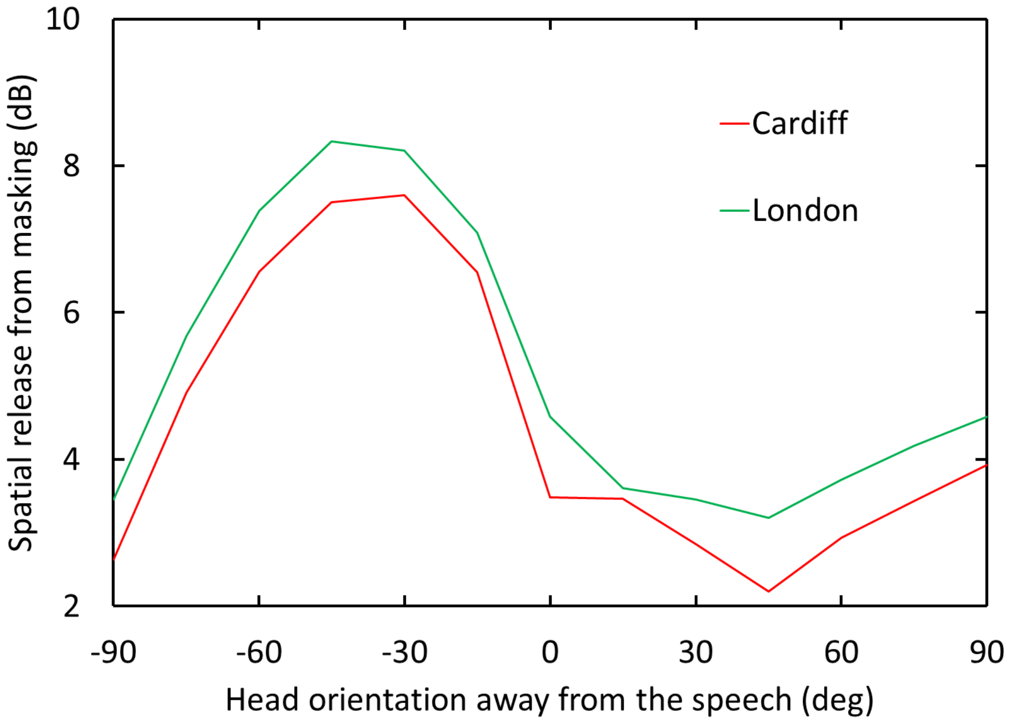


Figure 5.4: Illustration of the acoustical match between testing sites: SRM predictions for bilateral CI users in the T_0M_{90} configuration

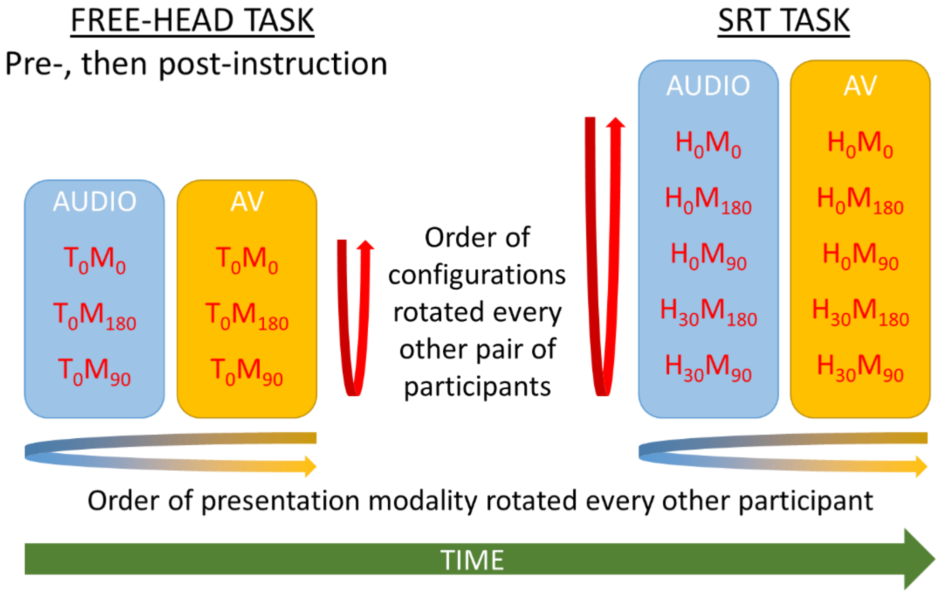


Figure 5.5: Schematic of task sequencing and condition rotation

Given the large number of conditions planned for CI users, there was only sufficient material in the SPIN corpus to perform two SRT repeat runs. As early results showed a lot of variability between repeat runs, it was decided to re-test all participants with the audio-only SRT protocol detailed in Section 3.3.4, using sentences from the IEEE corpus.

The resulting audio-only SRTs were analysed separately. The larger amount of material available (DA and CW speakers) and re-testing in the audio-only modality alone enabled 5-6 repeats in each condition. We expected the averaging of SRTs across the repeats to provide a reduced SE and hence a more precise measure of SF-SRM and HOB. The protocol was adapted by limiting the number of presentations of the first sentence to 4 and triggering the adaptive phase as soon as CI users recognised one or more of the five key words of the first sentence. We hereafter refer to this protocol as the ‘audio-only protocol’.

SRT data reliability was assessed as satisfactory for all retained participants by verifying that the word recognition averaged from the fourth staircase presentation onwards had a stable mean and a standard deviation not exceeding 10% for the trials that had an impact on reported outcomes.

5.3. Free-head-listening experimental outcomes

As in Chapter 4, all head-orientation tracks are plotted here relative to the SNR reached in the collocated runs within the same presentation modality. They are therefore plotted as a function of SRM and in the context of SRM predictions, with a subjective SRM highlighted at the end of each track and the corresponding final head orientation averaged over the last 10 seconds of a track.

5.3.1. Age-matched normally hearing listeners

5.3.1.1. Spontaneous head orientations

Figures 5.6 and 5.7 show example tracks in the T_0M_{180} and T_0M_{90} configurations respectively. Age-matched NH listeners were found to spontaneously make use of head orientation in only 30% of trials overall. However, 6 out of 10 listeners made use of head orientation at least once during the undirected runs. There was no indication that listeners preferentially turned one way rather than the other in the symmetrical, T_0M_{180} configuration. In contrast, 4 out of 10 listeners turned toward the masker (the correct direction to obtain a HOB) at T_0M_{90} in at least one presentation modality when only one listener ever turned the wrong way (in the audio-only modality).

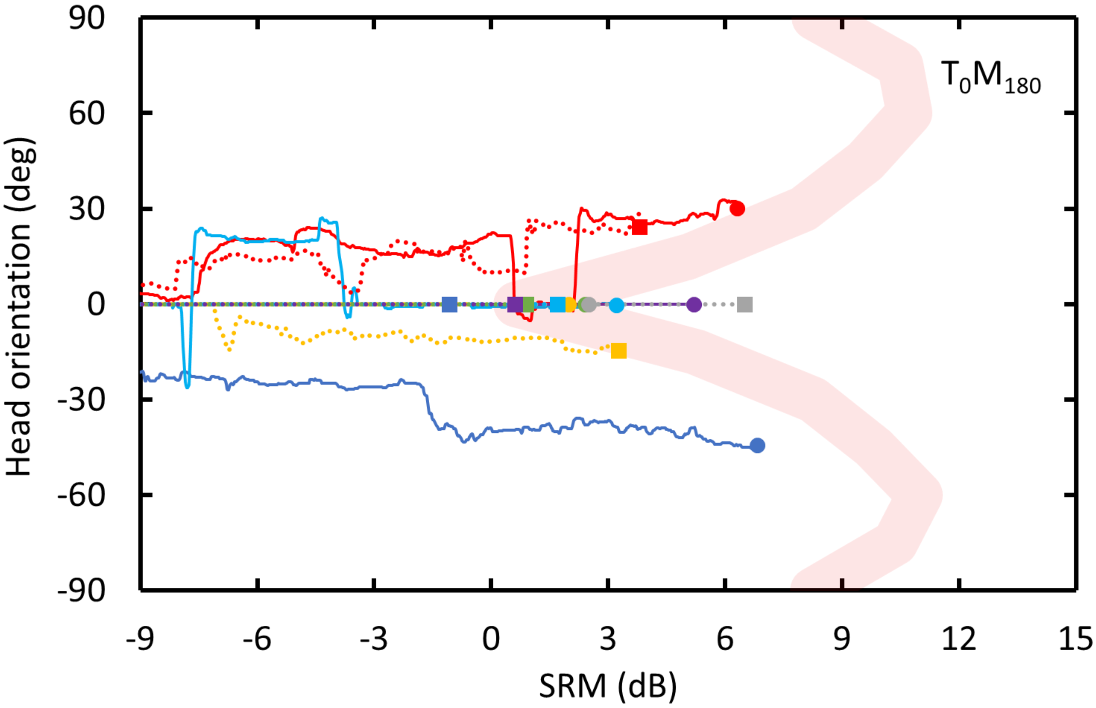


Figure 5.6: T_0M_{180} age-matched NH listeners' audio/AV pre-instruction head tracks (solid/dotted lines), subjective SRMs (circles/squares) and predictions (pink bands). One line and symbol colour is used per listener

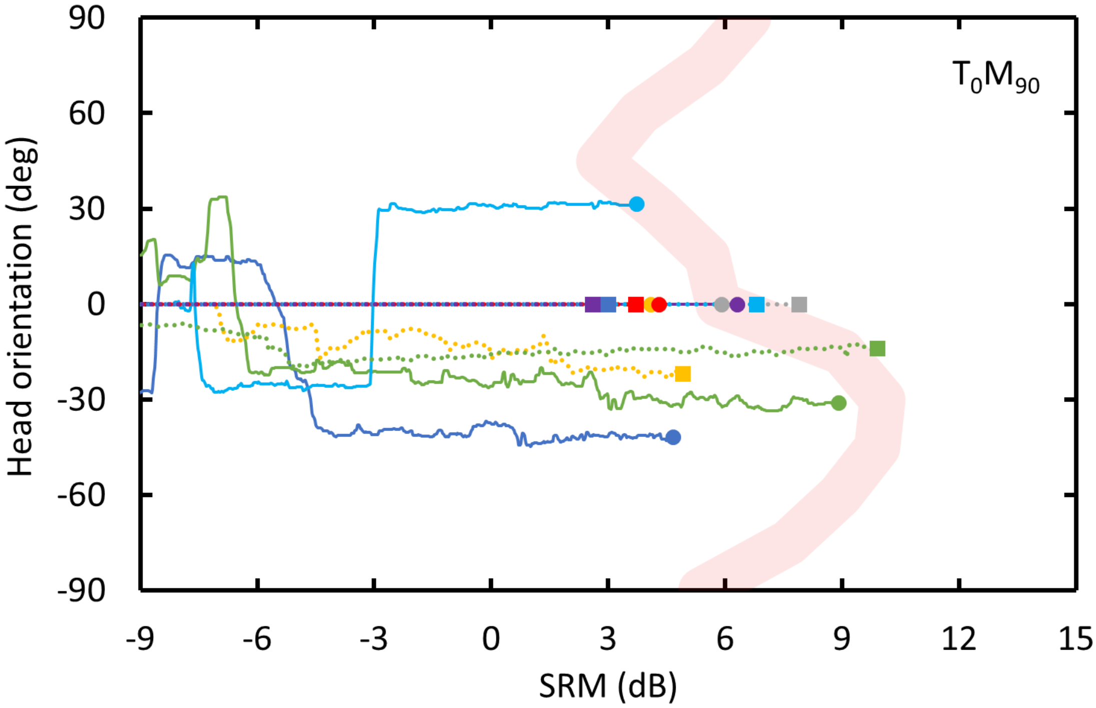


Figure 5.7: T_0M_{90} age-matched NH listeners' audio/AV pre-instruction head tracks (solid/dotted lines), subjective SRMs (circles/squares) and predictions (pink bands). One line and symbol colour is used per listener

Although AV presentation did not significantly reduce the amount of head orientation away from the speech ($F(1,9) = 3.88$, $p = 0.08$), those who ventured away from facing the speech did not exceed 25° when the speaker was visible, but typically exceeded 30° in audio-alone. This may be an indication that listeners were spontaneously lip-reading during AV presentation. Overall, none of the listeners went beyond 45°.

5.3.1.2. Effect of instruction on head orientations

Figure 5.8 and 5.9 plot the post-instruction head-tracks in the T_0M_{180} and T_0M_{90} configurations respectively. As with young adult listeners, instruction gave rise to an increase in the amount of head orientation away from the speech ($F(1,9) = 68.61$, $p < 0.001$). Overall, AV presentation significantly reduced the amount of head orientation ($F(1,9) = 27.28$, $p < 0.002$).

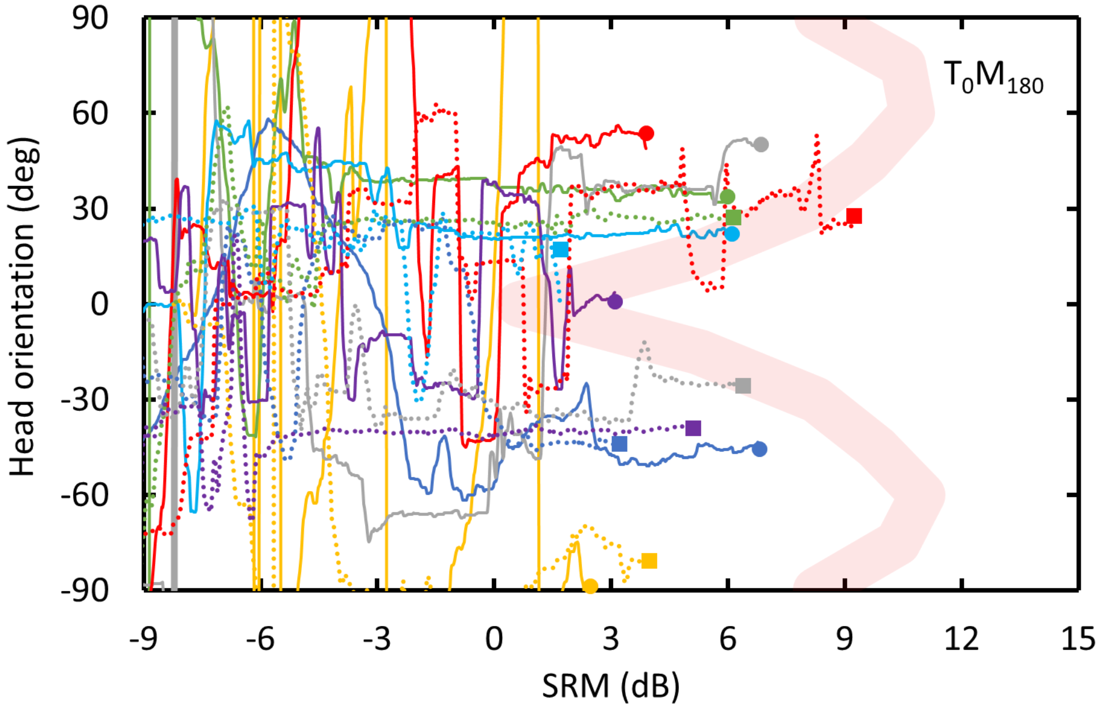


Figure 5.8: T_0M_{180} age-matched NH listeners' audio/AV post-instruction head tracks (solid/dotted lines), subjective SRMs (circles/squares) and predictions (pink bands). One line and symbol colour is used per listener

Post-instruction alone, AV presentation reduced the amount of head orientation from 71° to 37°, as 8 out of 10 listeners chose to settle at head orientations below 45°, probably to stay within lip-reading range. A significant interaction was found between instruction status and presentation modality (F(2,18) = 28.40, $p < 0.001$), indicating that AV presentation reduced the effect of instruction on head turns.

Post-instruction, listeners did not preferentially turn one way rather than the other in the T_0M_{180} configuration. Contrasting with pre-instruction findings, a majority of listeners chose to turn away from the masker in the T_0M_{90} configuration, as though to get away from the masker and even though this was the incorrect way to gain a HOB.

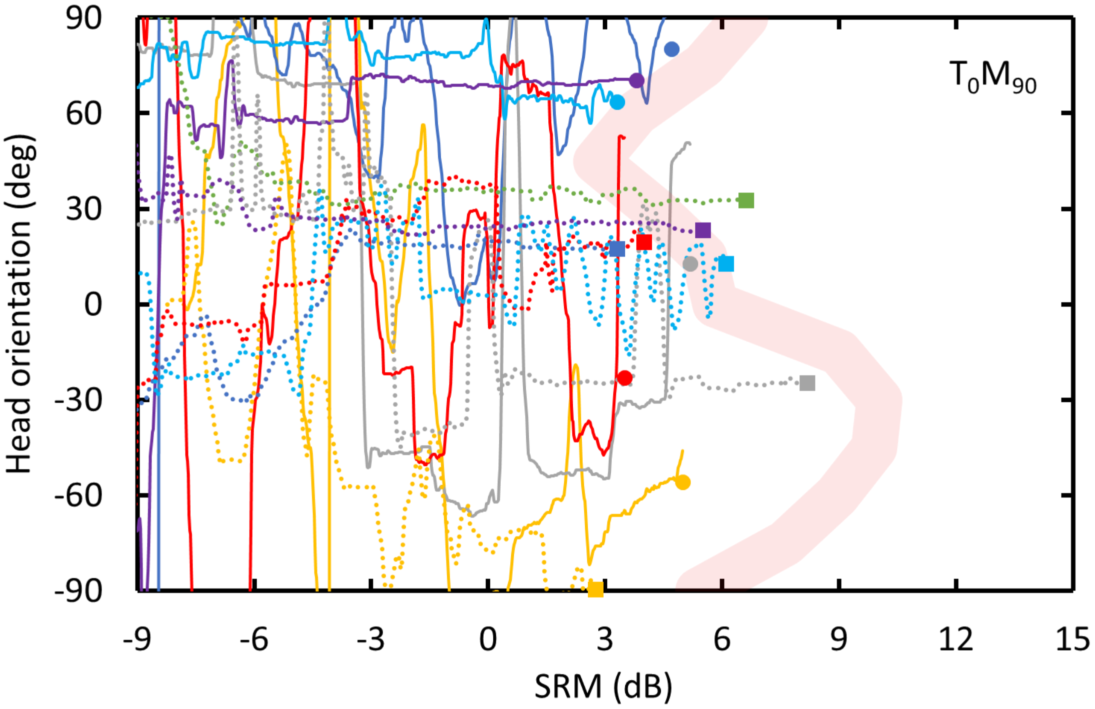


Figure 5.9: T_0M_{90} age-matched NH listeners' audio/AV post-instruction head tracks (solid/dotted lines), subjective SRMs (circles/squares) and predictions (pink bands). One line and symbol colour is used per listener

5.3.1.3. Subjective SRM vs. model predictions at final head orientations

Figure 5.10 presents the subjective SRM data next to cumulative model predictions for SF-SRM and maximum HOB (gained at optimum head orientations). As a result of limited spontaneous head movements, subjective SRMs of only 2.6 and 4.4 dB were reached in the T_0M_{180} and T_0M_{90} configurations respectively. This was markedly less than BRIR-model predictions of 11.0 and 10.0 dB for an optimum head orientation. As for young adult listeners, subjective SRM was positively correlated with the listeners' final unsigned head orientation in the T_0M_{180} configuration (r = 0.63, F(1,18) = 11.70, $p <$ 0.005), indicating that this benefit is related to orienting away from the speech source.

Subjective SRMs significantly increased as a result of instruction and increased head turns, but only by 0.8 dB (F(1,9) = 11.05, $p < 0.01$) with 4.5 and 4.1 dB reached in the T_0M_{180} and T_0M_{90} configurations respectively. The small additional benefit of head orientation suggests that older adults made even less effective use of head orientation than younger adults. Although they improved at T_0M_{180}, they did worse at T_0M_{90} and that was due to more than half of the participants turning away from the masker, in a direction that led to a negative SRM.

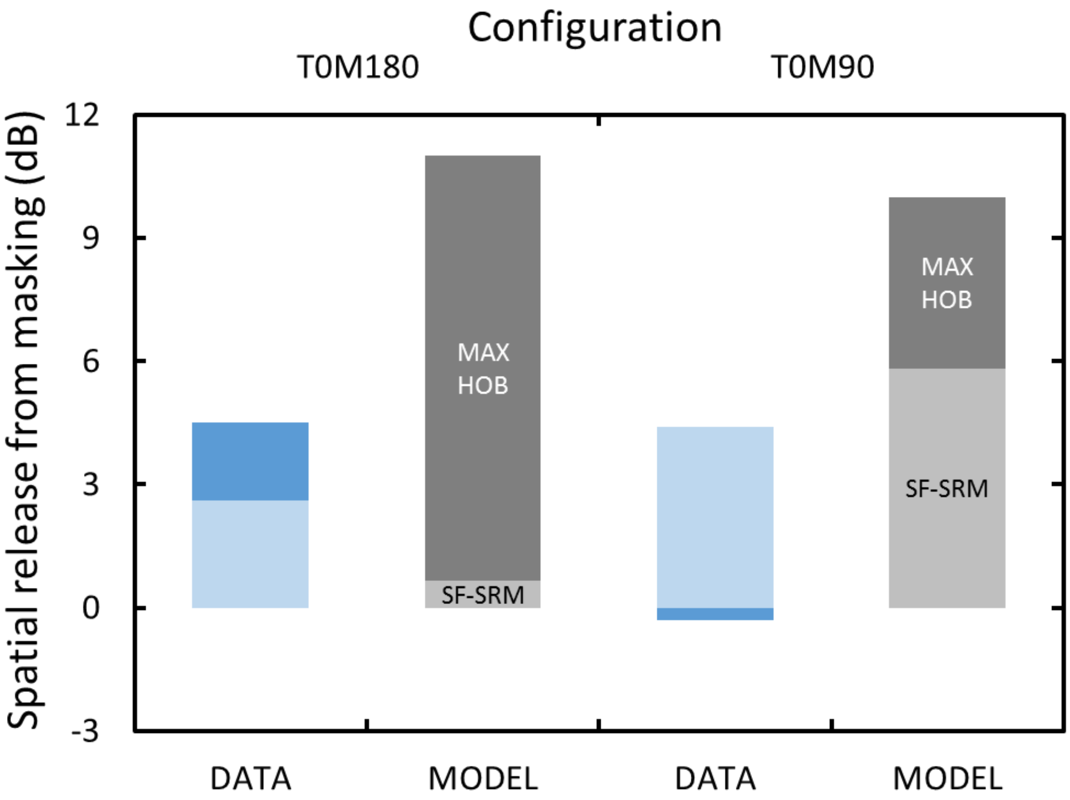


Figure 5.10: Age-matched NH listeners' subjective SRM gained pre- (pale bars) and additionally post-instruction (dark bars) in the T_0M_{180} and T_0M_{90} configurations. SRM model predictions for cumulative SF-SRM and maximum HOB

Figure 5.11 combines pre- and post-instruction SRM data points and plots them against the predicted SRM for final head orientations at T_0M_{180} and T_0M_{90}. As for young NH listeners, to remove any variability relating to lip-reading, only the audio data is presented here. A significant correlation was found between subjective SRM data and predictions at T_0M_{180} ($r = 0.58$, $t\,(18) = 3.00$, $p < 0.01$, slope 0.29 with 0.10 SE). However, no correlation was found in the T_0M_{90} configuration, presumably due to small or incorrect head turns combined with the high variability of the data.

Combining the two spatial configurations led to an overall significant correlation between subjective SRMs and predictions ($r = 0.35$, $t\,(38) = 2.30$, $p < 0.03$) with a 0.23 slope (0.10 SE), as illustrated in Figure 5.12. Clearly, as age-matched NH listener's head orientations were suboptimal, the data variability makes the regression slope unreliable, but the objective SRM data (Section 5.4) may shed some light as to why.

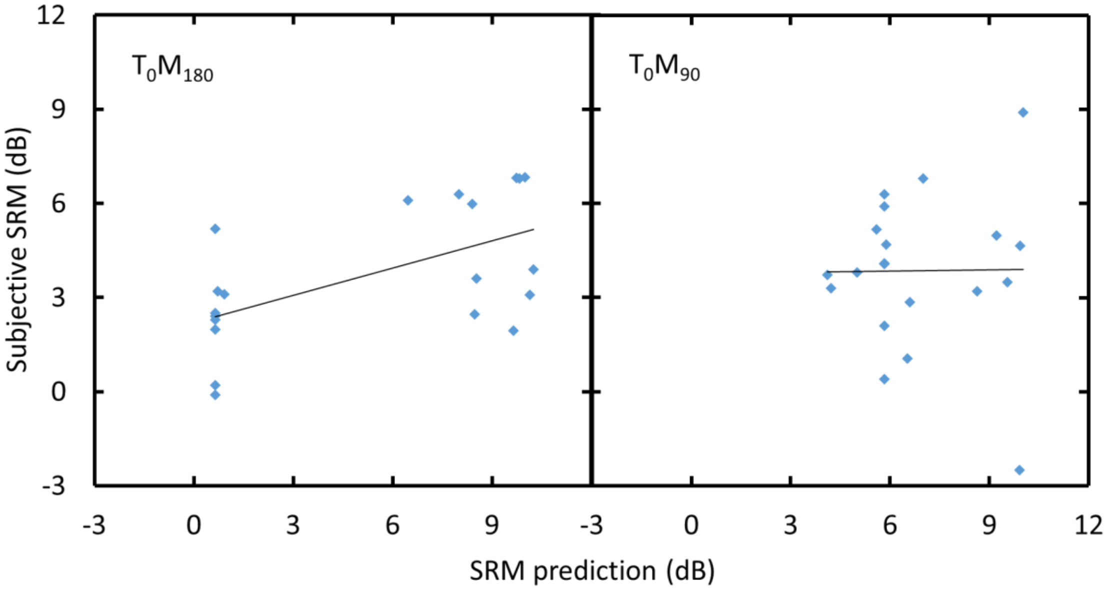


Figure 5.11: Correlation between age-matched NH listeners' subjective SRMs and SRMs predicted for final head orientations in the T_0M_{180} and T_0M_{90} configurations

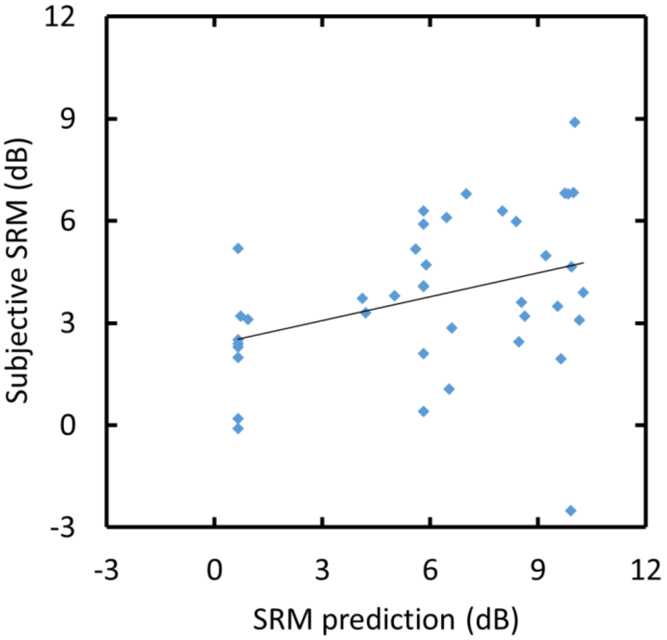


Figure 5.12: Correlation between age-matched NH listeners' subjective SRMs and SRMs predicted for final head orientations across configurations

5.3.2. CI users

5.3.2.1. Bilateral CI users' spontaneous head orientations

Figures 5.13 and 5.14 respectively show all tracks in the T_0M_{180} and T_0M_{90} configurations. Bilateral CI users were found to spontaneously turn their heads in only 25% of trials overall. Only 2 out of 8 participants ventured away from facing the AV speech, whilst never moving to head orientations that precluded lip-reading. In contrast, 5 out of 8 participants moved away from facing the speech in audio-alone.

5.3.2.2. Effect of instruction on head orientations

Figure 5.15 and 5.16 plot the post-instruction head-tracks in the T_0M_{180} and T_0M_{90} configurations respectively. As with NH listeners, instruction gave rise to an increase in head turns ($F(1,7) = 18.28$, $p < 0.005$). There was also a significant overall reduction in head turns in the AV modality ($F(1,7) = 21.95$, $p < 0.002$). Post-instruction alone, 6 out of 8 participants chose to stay within a range of head orientations that enabled lip-reading when the speaker was visible. This led to the mean head orientation away from the speaker being significantly reduced from 50° in audio-alone to 30° in AV ($F(1,7) = 15.78$, $p < 0.005$). This reflects how CI users are dependent on lip-reading when attending to speech.

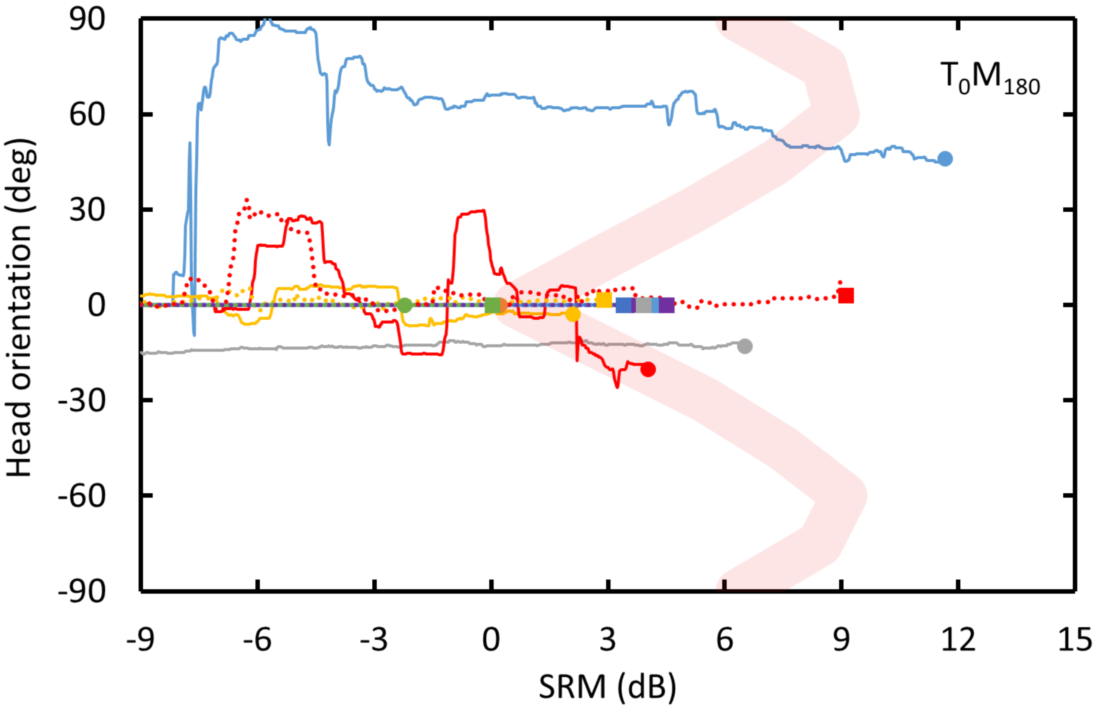


Figure 5.13: T_0M_{180} bilateral CI users' audio/AV pre-instruction head tracks (solid/dotted lines), subjective SRMs (circles/squares) and predictions (pink bands)

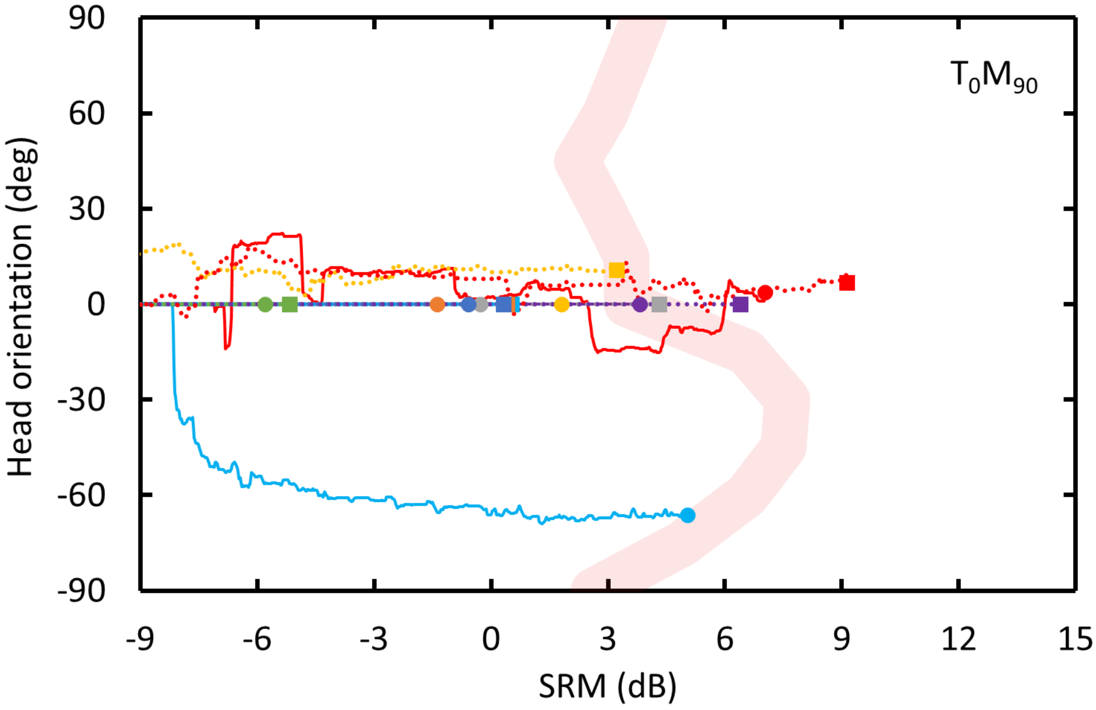


Figure 5.14: T_0M_{90} bilateral CI users' audio/AV pre-instruction head tracks (solid/dotted lines), subjective SRMs (circles/squares) and predictions (pink bands)

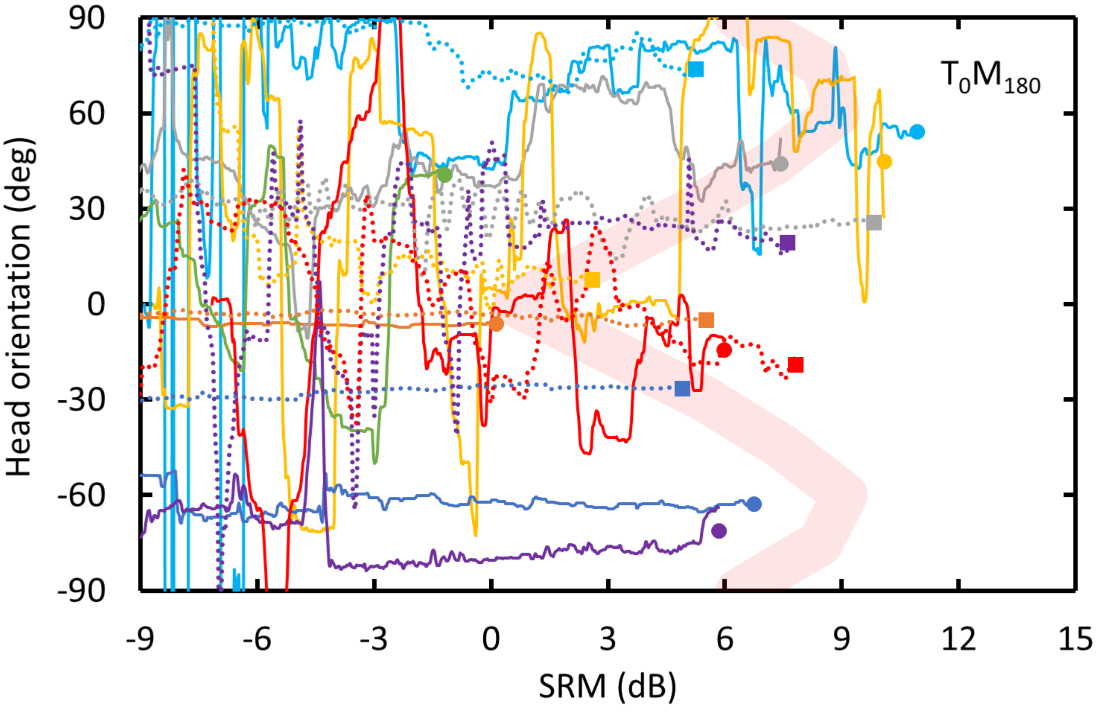


Figure 5.15: T_0M_{180} bilateral CI users' audio/AV post-instruction head tracks (solid/dotted lines), subjective SRMs (circles/squares) and predictions (pink bands)

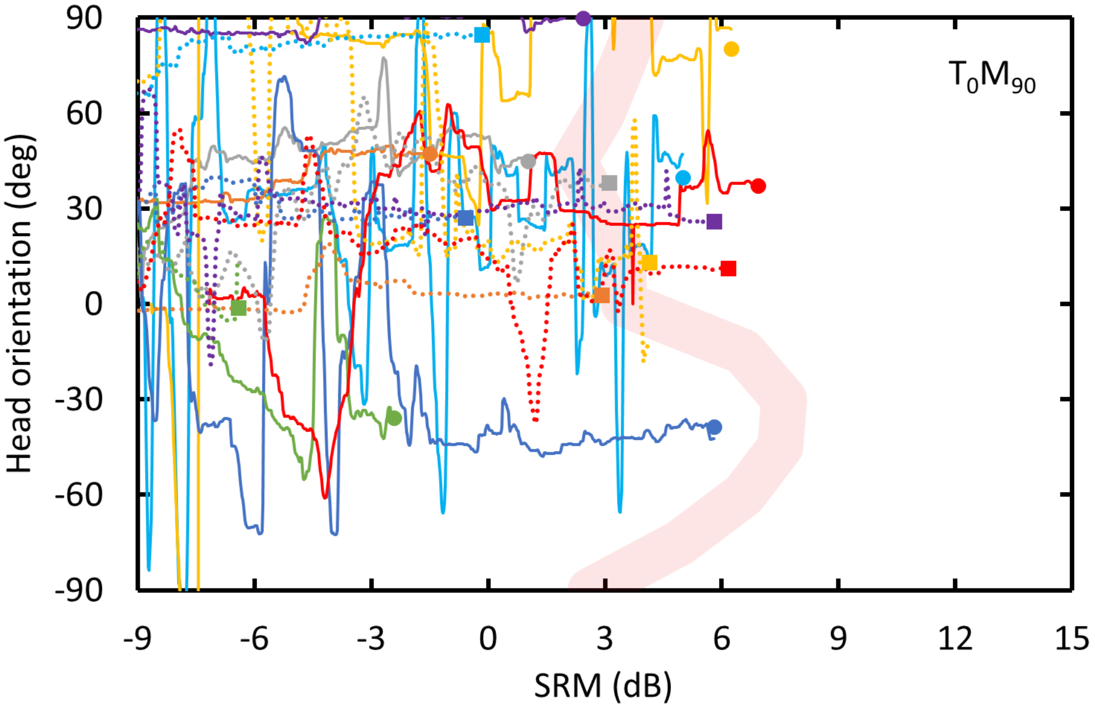


Figure 5.16: T_0M_{90} bilateral CI users' audio/AV post-instruction head tracks (solid/dotted lines), subjective SRMs (circles/squares) and predictions (pink bands)

5.3.2.3. Unilateral CI users' spontaneous head orientations

Figures 5.17 and 5.18 show all head-tracks in the T_0M_{180} and T_0M_{90} configurations respectively. Unilateral CI users were found to spontaneously make use of head orientation in only 10% of trials overall. Whilst (only) 2 out of 8 participants moved their heads in audio-only, AV presentation totally eradicated spontaneous head turns.

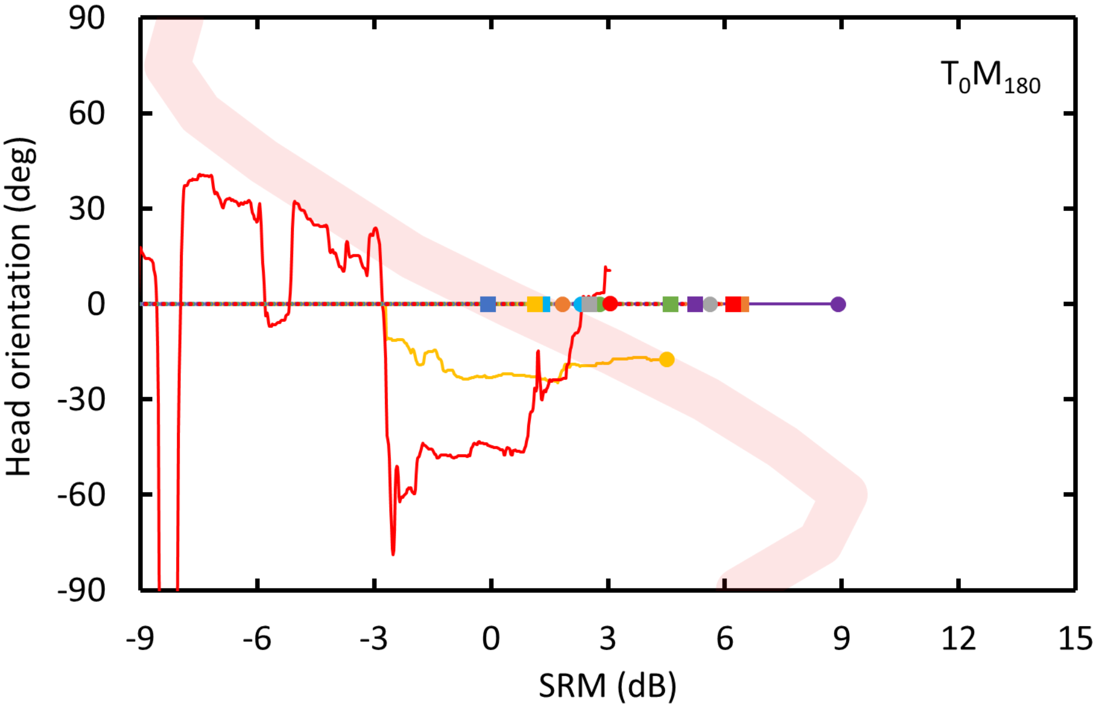


Figure 5.17: T_0M_{180} unilateral CI users' audio/AV pre-instruction head tracks (solid/dotted lines), subjective SRMs (circles/squares) and predictions (pink bands)

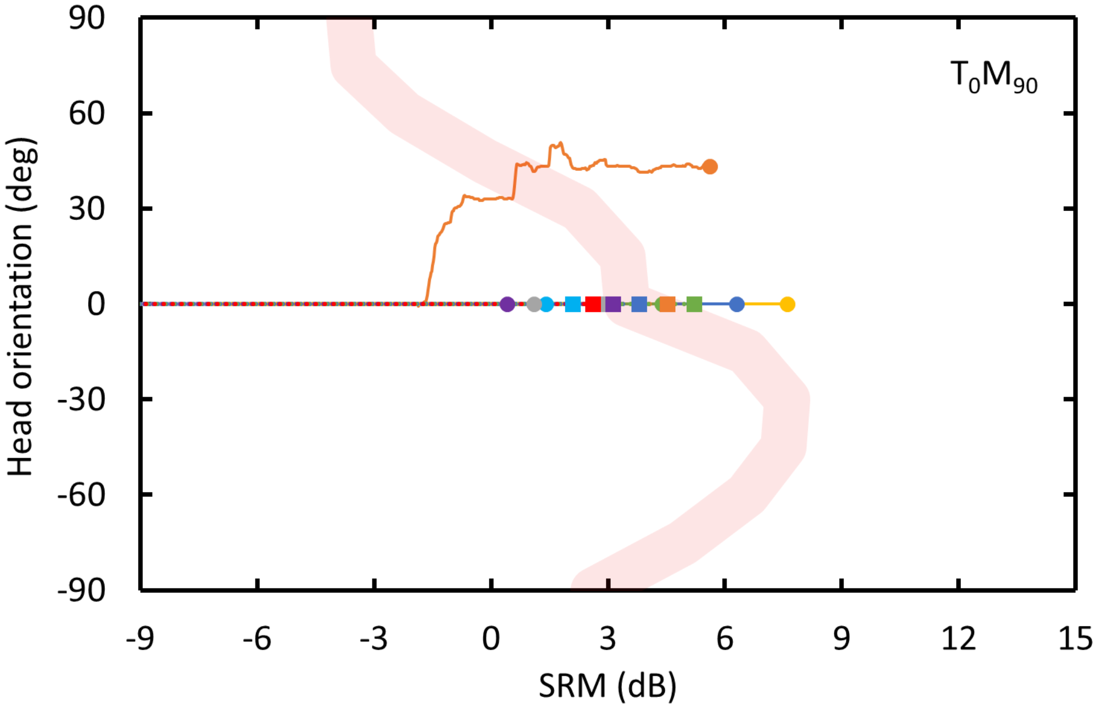


Figure 5.18: T_0M_{90} unilateral CI users' audio/AV pre-instruction head tracks (solid/dotted lines), subjective SRMs (circles/squares) and predictions (pink bands)

5.3.2.4. Effect of instruction on head orientations

Figures 5.19 and 5.20 plot the post-instruction head-tracks in the T_0M_{180} and T_0M_{90} configurations, respectively. As with all the other listeners, instruction gave rise to an increase in the amount of head orientation away from the speech ($F(1,7) = 51.42$, $p < 0.001$). Overall, AV presentation significantly reduced the amount of head turning ($F(1,7) = 15.81$, $p < 0.005$). Post-instruction and in the AV modality, 6 out of 8 participants kept their head orientation within a range compatible with lip-reading. In audio-only, 5 participants chose to explore head orientations beyond that range. As a result, AV presentation reduced the mean head orientation away from the speech from 57º to 30º. These results are very close to those obtained with bilateral CI users and reflect unilateral CI users' high reliance on lip-reading.

In term of direction of head turns, unilateral CI users almost unanimously opted to turn in the direction that brought their CI closer to the target speech, direction that the model predicted would lead to a HOB in both the T_0M_{180} and T_0M_{90} configurations. We cannot infer from the data as to why they almost always chose the correct direction when a majority of bilateral users and age-matched listeners turned the wrong way in the T_0M_{90} configuration. However, plausible reasons are that having a single CI removes the ambiguity of which way is best to turn. It also removes the reasons a majority of bilateral users and age-matched listeners may have had to move away from the masker direction, as though to get away from it.

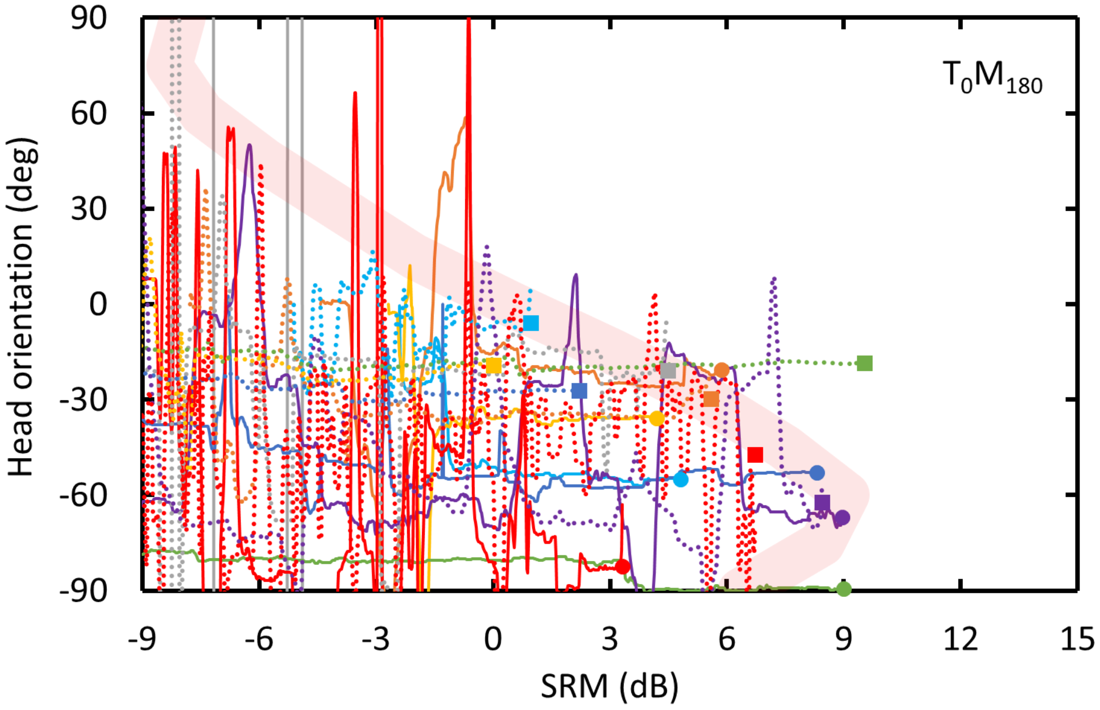


Figure 5.19: T0M180 unilateral CI users' audio/AV post-instruction head tracks (solid/dotted lines), subjective SRMs (circles/squares) and predictions (pink bands)

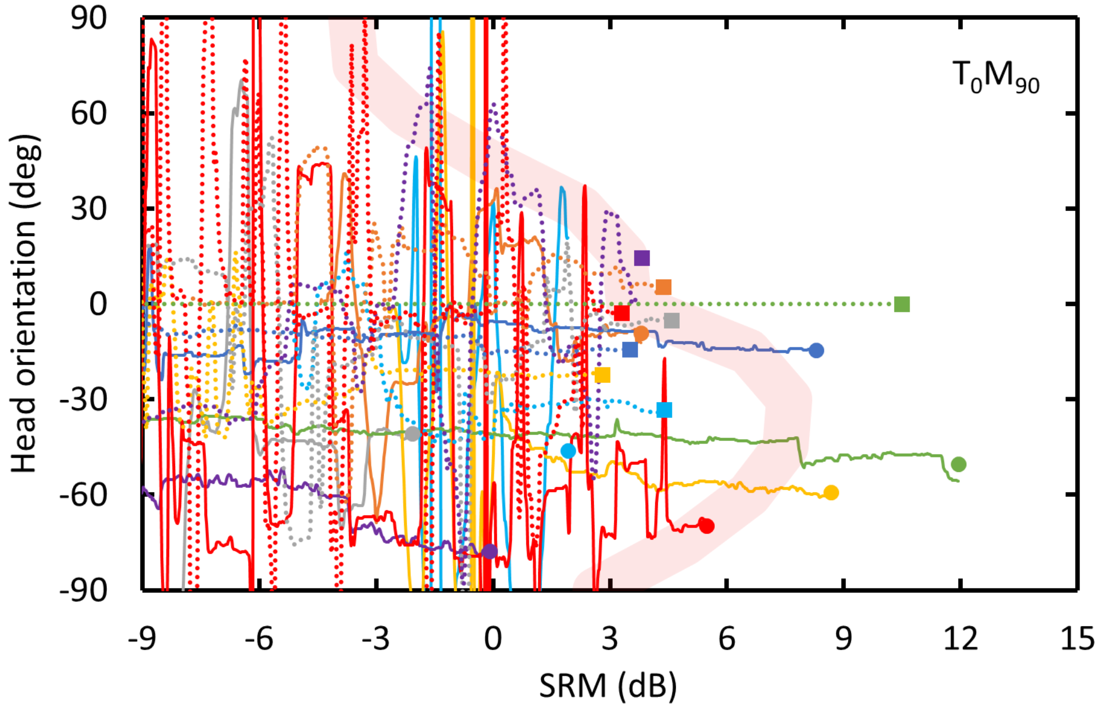


Figure 5.20: T$_0$M$_{90}$ unilateral CI users' audio/AV post-instruction head tracks (solid/dotted lines), subjective SRMs (circles/squares) and predictions (pink bands)

5.3.2.5. Subjective SRM vs. model predictions at final head orientations

Pre- and post-instruction mean subjective SRMs are displayed in Figure 5.21 next to SRM predictions for optimum head orientation. Sub-optimal head orientations pre-instruction led to subjective SRM reaching only 3.7 and 2.4 dB in T_0M_{180} and T_0M_{90} respectively. Subjective SRMs were small compared to the predicted 9.1 and 7.6 dB for optimal head orientations. Post-instruction and with increased head movements, CI users' subjective SRMs significantly increased on average by 1.2 dB ($F(1,7) = 5.06$, $p < 0.05$) to 4.9 and 3.6 dB. An immediate benefit to CI users' intelligibility of speech in noise was thus observed from a simple instruction to explore the benefit of head orientation. There was no CI user group effect on subjective SRM outcomes, nor any interaction between instruction and CI user group. The post-instruction outcomes remained suboptimal. This suggests that CI users could obtain a lot more HOB if given more specific guidance.

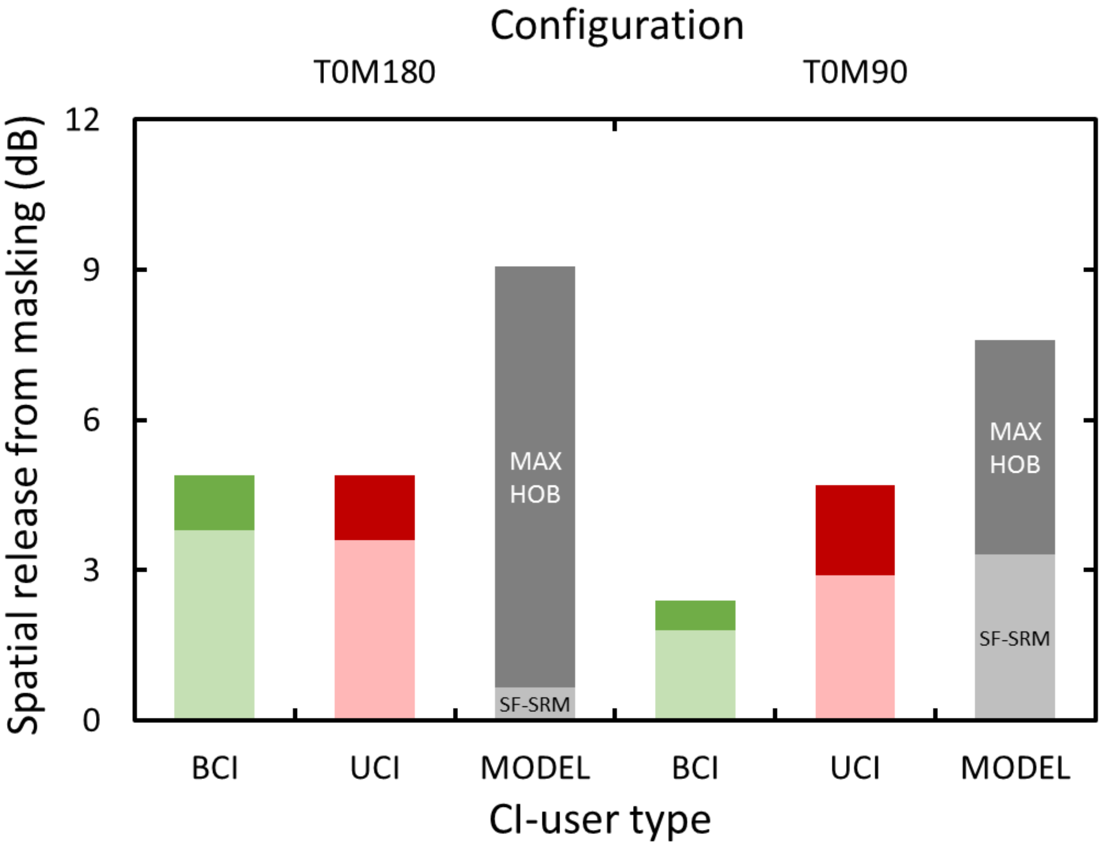


Figure 5.21: Bilateral (BCI) and unilateral (UCI) CI users' subjective SRM gained pre- (pale bars) and additionally post-instruction (dark bars) in the T_0M_{180} and T_0M_{90} configurations. Model predictions for SF-SRM and maximum HOB

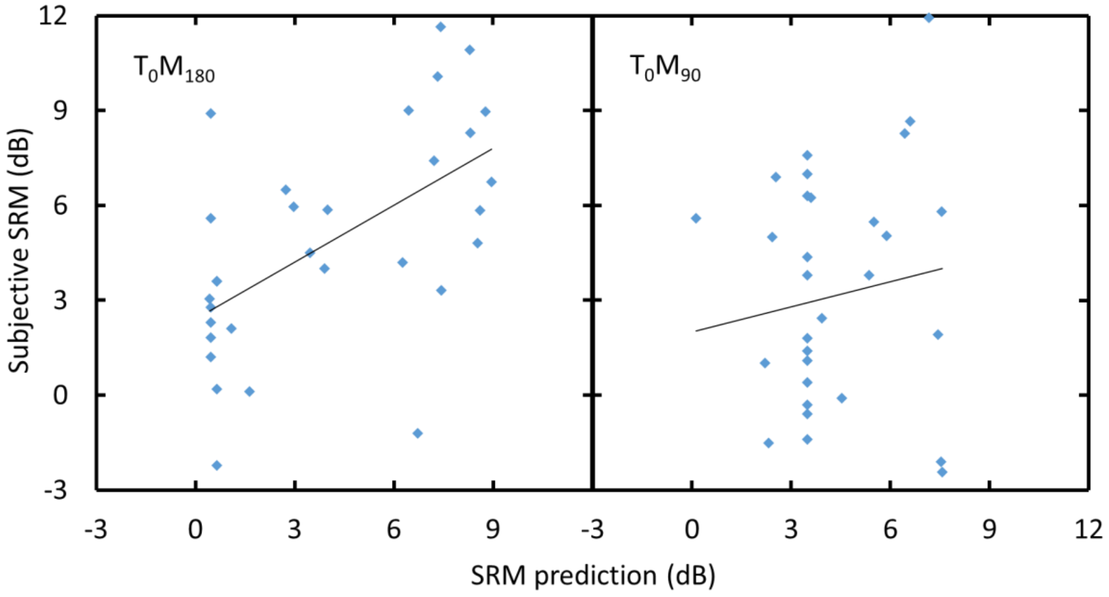


Figure 5.22: Correlation between CI users' subjective SRMs and SRMs predicted for final head orientations in the T_0M_{180} and T_0M_{90} configurations

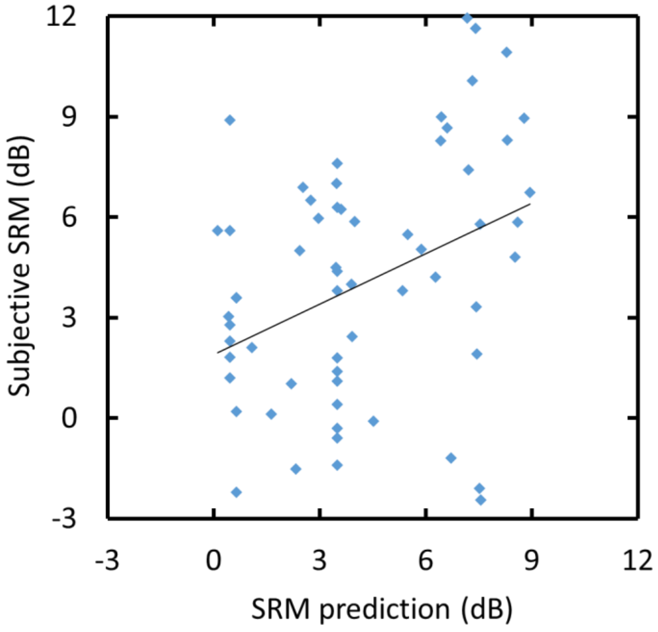


Figure 5.23: Correlation between CI users' subjective SRMs and SRMs predicted for final head orientations across configurations

As with age-matched NH listeners, a significant correlation was found between audio-only subjective SRM data and predictions for final head orientations at T_0M_{180} ($r = 0.57$, t (29) = 3.78, $p < 0.001$, slope 0.60) whilst no significant correlation was found in the T_0M_{90} configuration. Scatter plots are shown in Figure 5.22. As with NH listeners, the lack of correlation at T_0M_{90} was plausibly due to small or incorrect head turns combined with high data variability. The two spatial configurations, combined in Figure 5.23, led to an overall significant correlation between subjective SRMs and predictions (r = 0.36, t (60) = 2.96, p < 0.005) with a 0.50 slope (0.17 SE).

5.3.3. Comparisons between listener groups and overall effects

We consistently found across listener groups, that the amount of head orientation away from the speaker, whilst reduced by the speaker being visible, increased after instructions. With an increase in head turns, subjective SRM increased significantly post-instruction for both CI users and age-matched HN listeners. However, the SRM reached by both groups was small compared to that available to them according to the model.

An ANOVA for head orientation across all listener types and conditions confirmed that instruction significantly increased head turns (F(1,23) = 114.7, p < 0.001) and that AV modality significantly reduced them (F(1,23) = 58.93, p < 0.001). The significant interaction between instruction status and modality confirms that seeing the speaker reduced the effect of instruction (F(1,23) = 10.21, p < 0.001). There was however no significant effect of listener group found, nor any interaction between listener group and other factors. Limiting the analysis to the T_0M_{180} configuration provided the same result.

An ANOVA for subjective SRM across factors and listener groups confirmed that instruction did significantly improve performance by an average of 1.1 dB (F(1,23) = 9.68, p < 0.005), with no significant effect of listener group. The only interaction that was almost significant was between listener group and spatial configuration (F(2,23) = 3.31, p = 0.055), where bilateral CI users seemed, overall, to reach lower subjective SRMs at T_0M_{90} than at T_0M_{180}, in contrast with NH listeners and unilateral CI users. This will be considered later, together with objective measures of SRM.

5.4. Objective-SRT experimental outcomes

Quality assurance for the SRT data was established by computing the variability of the average number of correct words across runs. Given that SRTs were used for SRM calculation, it was most important that the variability of the means between runs be well controlled (rather than means stay close to 50%). Moreover, since the staircase methods employed were already extensively proven for NH listeners, comparison of variability between NH and both CI user groups was the most important quality assurance point. The NH across-runs maximum and mean SD of within-run percent-correct means were 12% and 8.5% respectively. These values were respectively larger by only 5% and 2% for CI users, which was considered proof that the adaptive tracks had adequately converged.

5.4.1. Age-matched normally hearing listeners

10 age-matched normally hearing listeners first performed 3 repeat runs with the SPIN sentences in the modified SRT protocol. The speech-facing SF-SRM and additional 30° HOB are plotted in Figure 5.24 for audio-alone and for AV presentations against model predictions for the two spatially separated configurations. The variability of outcomes is represented here, as in all figures of Section 5.4, by error bars of one standard error of the means (SE).

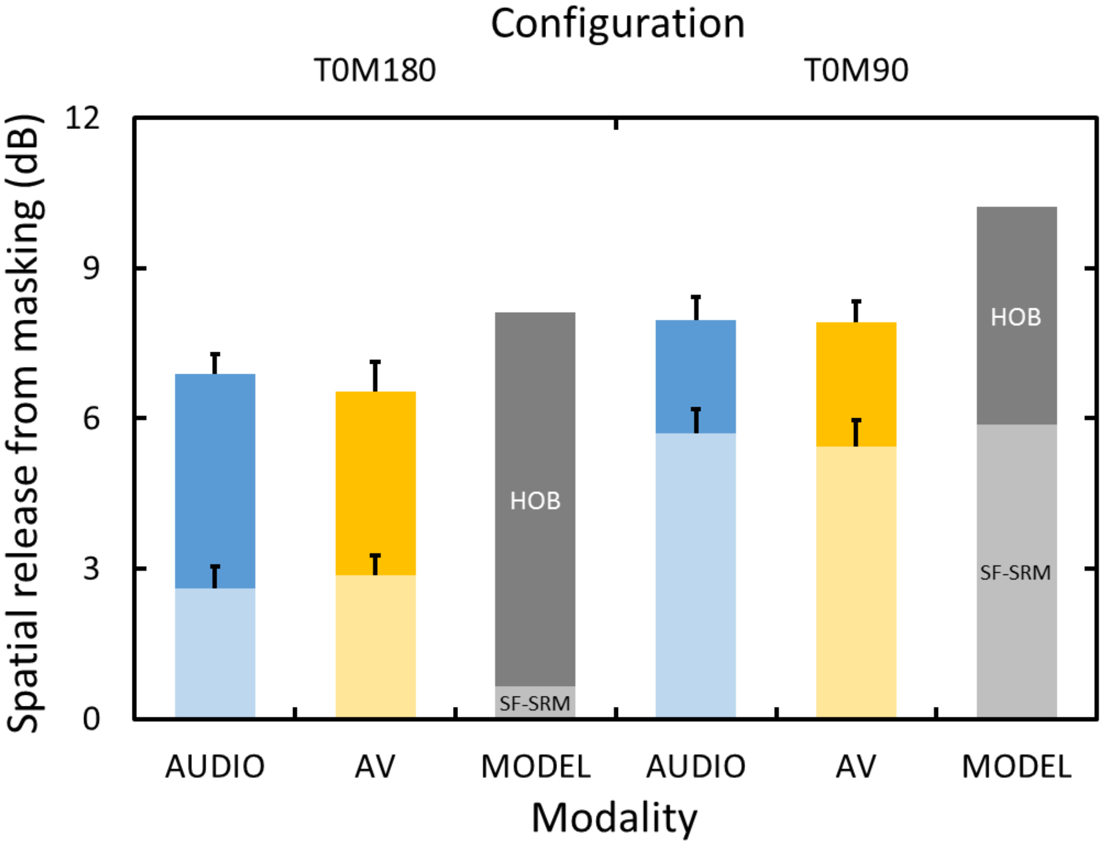


Figure 5.24: Age-matched NH listeners' SF-SRM (pale lower bars) and 30° HOB (dark upper bars) in audio and AV modalities against model predictions. Error bars are standard error of means

5.4.1.1. Speech-facing SRM

SF-SRM data in the T_0M_{180} and T_0M_{90} configurations are displayed as the pale lower bars of Figure 5.24 for audio alone and for AV presentation modalities.

In the T_0M_{180} configuration, SF-SRM was large (2.7 dB) compared to the BRIR-based prediction (0.7 dB), but almost identical to that found for young NH listeners. The difference between audio and AV SF-SRM was less than 0.2 dB, well within the standard errors (0.4 dB). For T_0M_{90}, audio and AV SF-SRMs were 5.7 and 5.4 dB (0.5 dB SE) respectively. As for young listeners, this compares well with the model prediction of 5.8 dB.

An ANOVA for SF-SRT confirmed that masker separation had a significant effect on SF-SRT (*F*(2,18) = 99.01, $p < 0.001$), with pairwise comparison between H_0M_0, H_0M_{180} and H_0M_{90} all showing significant differences in SRTs ($p < 0.001$).

5.4.1.2. Additional 30º HOB

The HOB obtained in T_0M_{180} and T_0M_{90} configurations can be seen in Figure 5.24 (dark upper bars), for audio-alone and for AV modalities. At T_0M_{180}, HOBs of 4.27 and 3.67 dB were obtained in audio and AV, respectively (with 0.4 and 0.6 dB SE). This compares to a model prediction of 7.5 dB. The reference for these benefits are the SRTs obtained at T_0M_{180} which we assumed were boosted by listeners' unintentional deviation from facing the speech as they focussed on the listening task. However, these HOBs were 1.1 dB lower than those of the younger listeners (see Chapter 4). At T_0M_{90}, HOBs of 2.3 and 2.5 dB were obtained in audio and AV, respectively (with 0.4 dB SE), compared to 4.4 dB predicted. Again, this was 0.9 dB lower than for younger adults. An ANOVA for SRM within each presentation modality confirmed that head orientation had a significant, beneficial effect (*F*(1,9) = 93.15, $p < 0.001$).

The cumulated SF-SRM and HOB in the T_0M_{180} configuration were 6.9 and 6.5 dB in audio and AV respectively, compared to the 8.1 dB model prediction. As with young adults, the fact that cumulated benefits almost reached model predictions is consistent with age-matched NH listeners having deviated somewhat from facing the speech at H_0M_{180}.

5.4.1.3. Lip-reading benefit

Figure 5.25 displays the age-matched NH listeners' LRB measured in each spatial configuration. The LRB ranged across participants and configurations from -0.3 to +4.6, averaging 2.8 dB (0.3 dB SE on average). This is in good agreement with the young adults' outcome of 3.0 dB (see Chapter 4). The cross-participants means ranged from 2.5 to 3.1 dB across configurations.

As with young adults, an ANOVA for SRTs across presentation modalities and configurations found LRB to be significant (*F*(1,9) = 547, $p < 0.001$). The absence of interaction between modality and configuration (*F*(4,36) = 0.41, $p > 0.8$) confirmed that configuration had no detrimental effect on LRB. Most importantly, LRB was unaffected by a 30º head turn (*F*(1,9) = 0.78, $p = 0.40$).

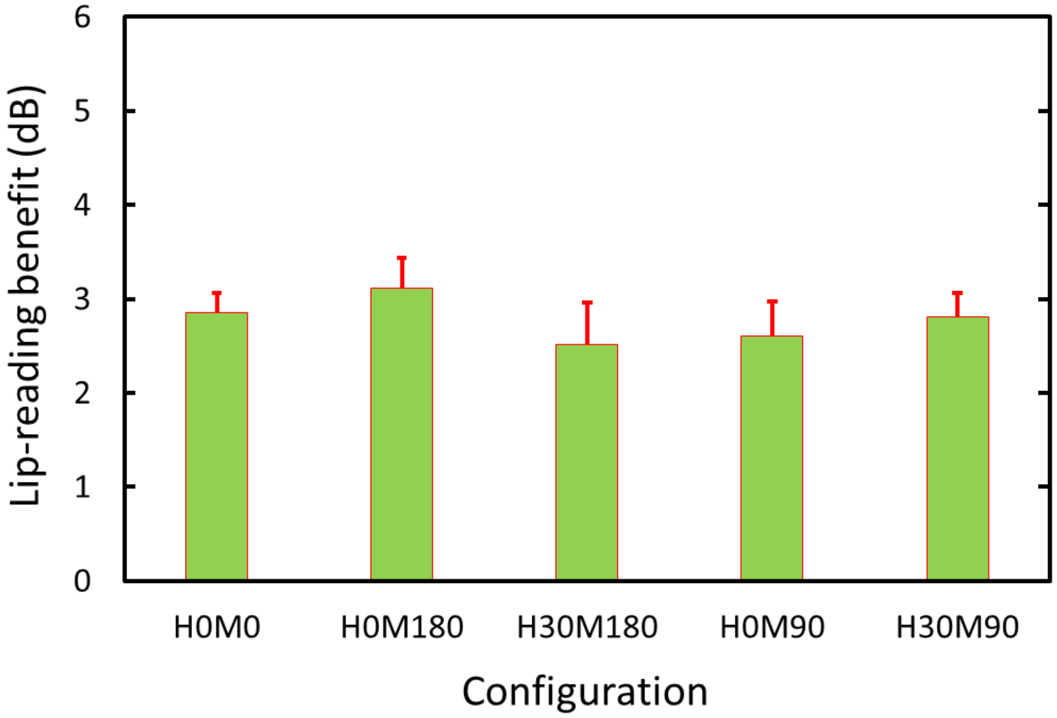


Figure 5.25: Age-matched NH listeners’ LRB, error bars are standard error of means

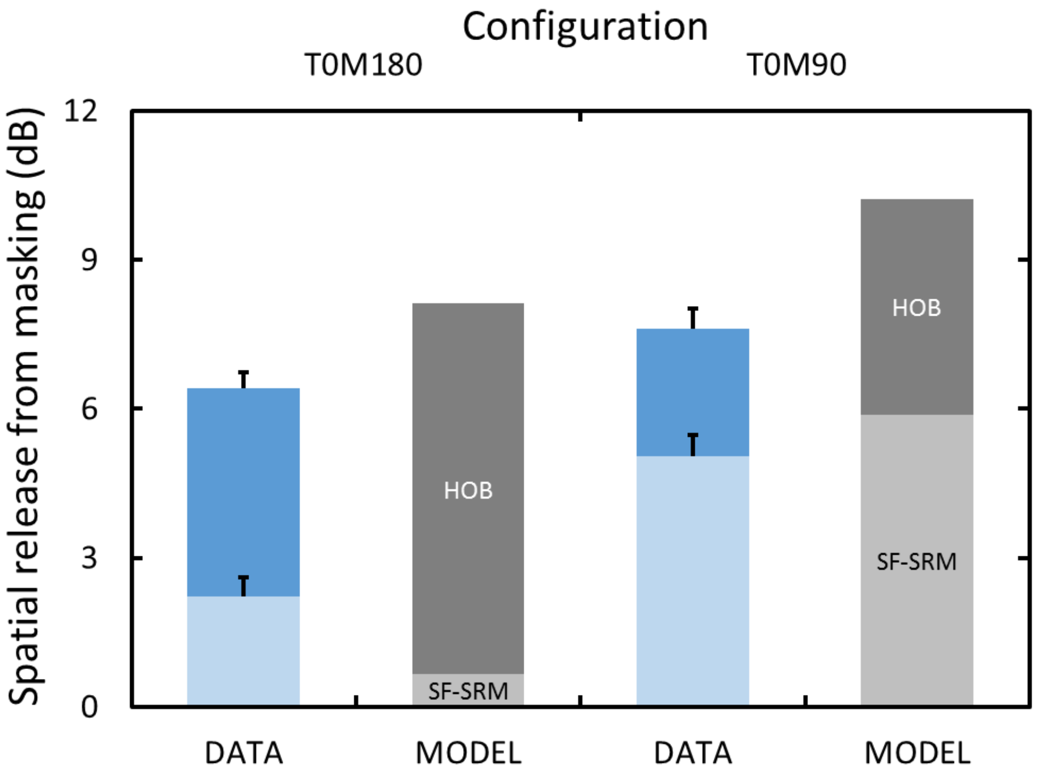


Figure 5.26: Age-matched NH listeners’ audio-only SF-SRM (pale lower bars) and 30º HOB (dark upper bars) against model predictions, error bars are standard error of means

5.4.1.4. Precise audio SRT measures

The same 10 age-matched normally hearing listeners performed 6 repeats with the audio-only protocol (IEEE sentences). The speech-facing SF-SRM and additional 30º HOB are plotted in Figure 5.26 against model predictions for the two spatially separated configurations. As expected, audio-only results obtained with the modified protocol were reproduced here within measurement error.

The SF-SRM was 2.2 and 5.1 dB and the 30º HOB was 4.2 and 2.6 dB at T_0M_{180} and T_0M_{90} respectively (0.4 dB average SE). An ANOVA for speech-facing SRT confirmed that masker separation had a significant effect on SF-SRT ($F(2,18) = 77.02$, $p < 0.001$), with pairwise comparisons between H_0M_0, H_0M_{180} and H_0M_{90} all showing significant differences in SRTs ($p < 0.001$). An ANOVA for SRM confirmed that head orientation had a significant, beneficial effect ($F(1,9) = 141.02$, $p < 0.001$).

The cumulated SF-SRM and HOB were 6.4 and 7.6 dB at T_0M_{180} and T_0M_{90} respectively, compared to the 8.1 and 10.2 dB model predictions. Young NH listeners' cumulated SRMs were 7.6 and 8.3 dB at T_0M_{180} and T_0M_{90}, respectively (see Chapter 4). The lower cumulated SRM of age-matched NH suggests an age-related loss of SRM. This will be discussed when comparing all outcomes.

5.4.2. Bilateral CI users

8 bilateral CI users first performed 3 repeat runs with the modified SRT protocol. The speech-facing SF-SRM and additional 30º HOB are plotted in Figure 5.27 for audio-alone and for AV presentations against model predictions for the two spatially separated configurations. Note here that model predictions are lower than for NH listeners since for bilateral CI users only the better-ear component of SRM is considered relevant.

5.4.2.1. Speech-facing SRM

The T_0M_{180} SF-SRM was large (2.9 dB) compared to the BRIR-based prediction (0.7 dB) and almost identical to that found for NH listeners. The difference between audio and AV SF-SRM is again very small (0.1 dB) and well within the standard errors (~1 dB). For T_0M_{90}, audio and AV SF-SRMs were 3.5 and 4.1 dB (1 dB SE) respectively. This compares with a model prediction of 3.3 dB.

An ANOVA for speech-facing SRT confirms that masker separation had a significant effect on SF-SRT ($F(2,14) = 16.62$, $p < 0.001$), with pairwise comparison between H_0M_0 and H_0M_{180} or H_0M_{90} all showing significant differences in SRTs ($p < 0.008$). Given the measurements' relatively large SE, the difference between H_0M_{180} and H_0M_{90} SRTs did not reach significance ($p = 0.09$).

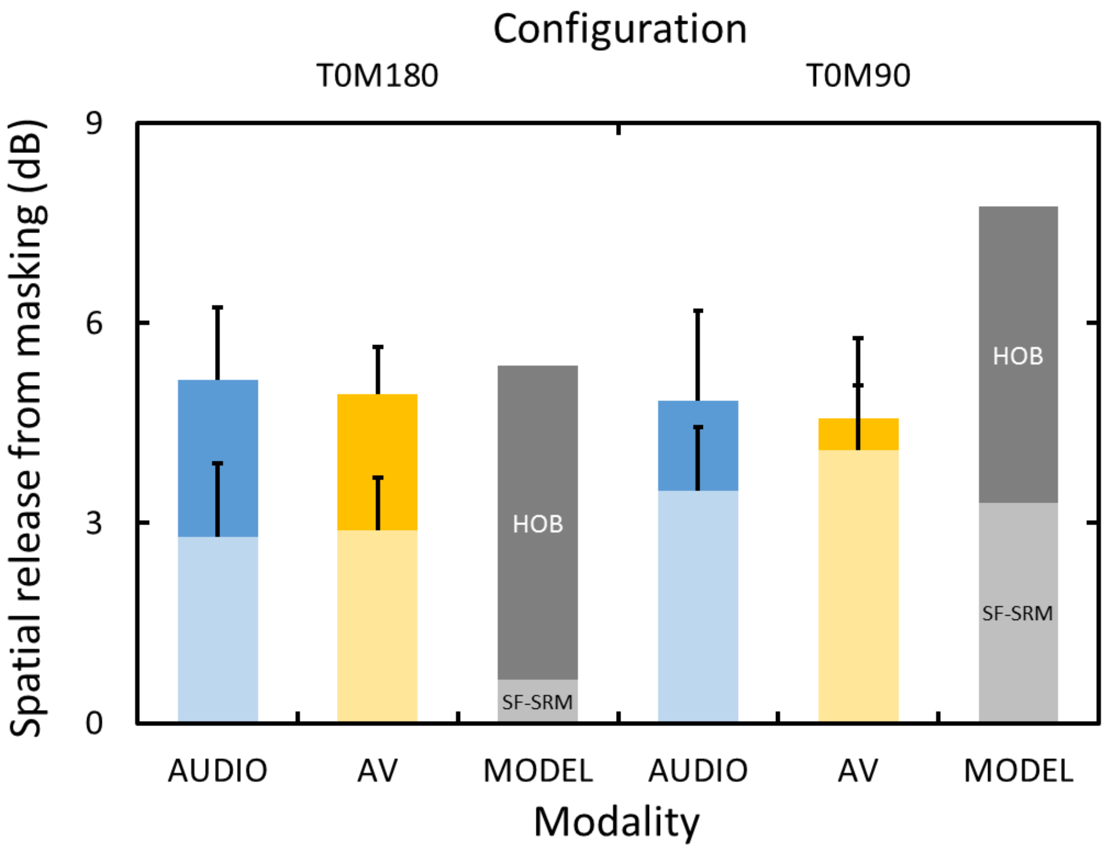


Figure 5.27: Bilateral CI users' SF-SRM (pale lower bars) and 30º HOB (dark upper bars) in audio and AV modalities against model predictions. Error bars are standard error of means

5.4.2.2. Additional 30º HOB

T_0M_{180} HOBs of 2.4 and 2.0 dB (1.1 and 0.7 dB SE) were observed in audio and AV, respectively. This was markedly lower than the 4.7 dB model prediction. T_0M_{90} HOBs of 1.3 and 0.5 dB (1.4 and 1.2 dB SE) were obtained in audio and AV respectively, compared to 4.4 dB predicted. An ANOVA for SRM confirmed that head orientation had a significant, beneficial effect ($F(1,7) = 27.87$, $p < 0.002$) with no significant interaction with spatial configuration ($F(1,7) = 1.90$, $p = 0.21$).

The cumulated SF-SRM and HOB in the T_0M_{180} configuration were 5.1 and 4.9 dB in audio and AV respectively, compared to the 5.4 dB model prediction. The fact that cumulated benefits almost reached model predictions suggests that bilateral CI users also deviated from facing the speech at H_0M_{180}.

5.4.2.3. Lip-reading benefit

Figure 5.28 displays the bilateral CI users' LRB measured in each spatial configuration. The benefit ranged across participants and configurations from -2 to 11 dB, averaging 4.3 dB (0.9 dB SE on average). This is significantly larger than the 2.8 dB found for the age-matched NH controls (see group comparison in Section 5.4.4.2). The LRB ranged from 4.1 to 4.9 dB across configurations and from 1.5 to 6.2 dB across participants.

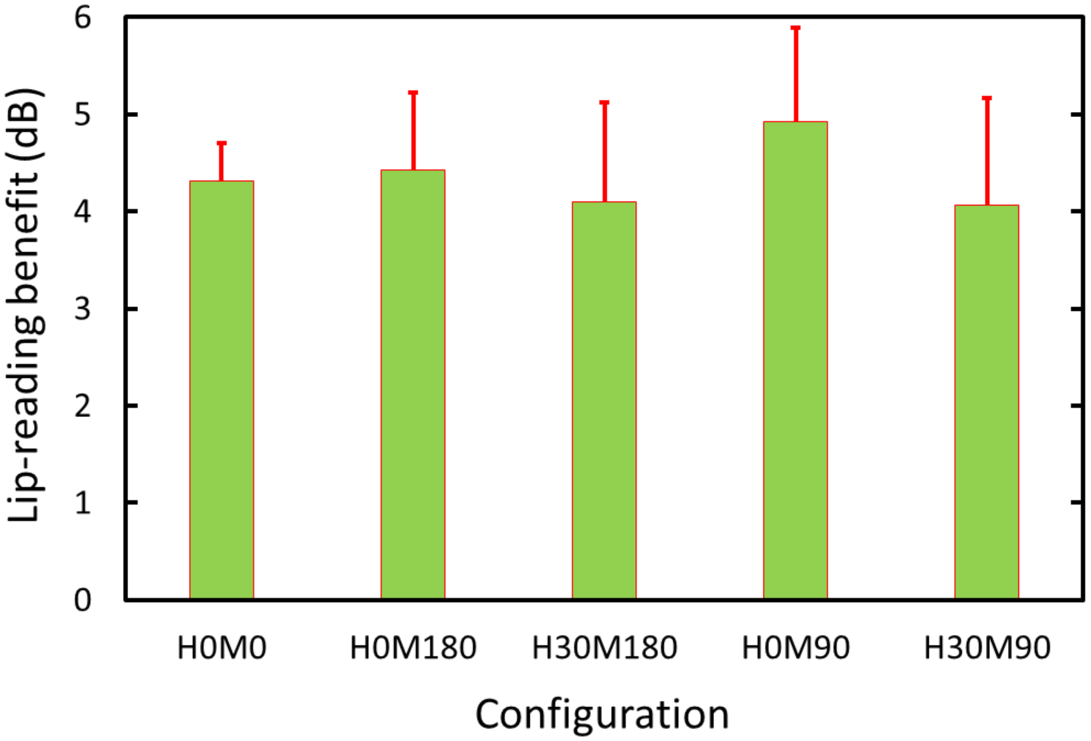


Figure 5.28: Bilateral CI users' LRB, error bars are standard error of means

As with NH listeners, an ANOVA for SRTs confirmed a significant benefit of visual cues ($F(1,7) = 79.36$, $p < 0.001$) with no interaction between modality and configuration ($F(4,28) = 0.11$, $p > 0.9$). LRB was specifically confirmed to be unaffected by a 30° head turn ($F(1,7) = 0.26$, $p = 0.62$).

5.4.2.4. Precise audio SRT measures

All 8 bilateral CI users performed 5-6 repeats with the audio-only protocol. The speech-facing SF-SRM and additional 30° HOB are plotted in Figure 5.29 and are consistent, within measurement error, with results obtained with the modified protocol.

SF-SRM was 2.9 and 3.1 dB at T_0M_{180} and $T_0M_{90,}$ respectively, and 30° HOB was 1.9 and 1.5 dB (0.55 dB average SE). An ANOVA for speech-facing SRT confirmed that masker separation had a significant effect on SF-SRT ($F(2,14) = 17.99$, $p < 0.001$), with pairwise comparison between H_0M_0 and H_0M_{180} or H_0M_{90} showing significant differences in SRTs and therefore significant SF-SRMs ($p < 0.002$). An ANOVA for SRM confirmed that head orientation had a significant, beneficial effect ($F(1,7) = 18.86$, $p < 0.003$).

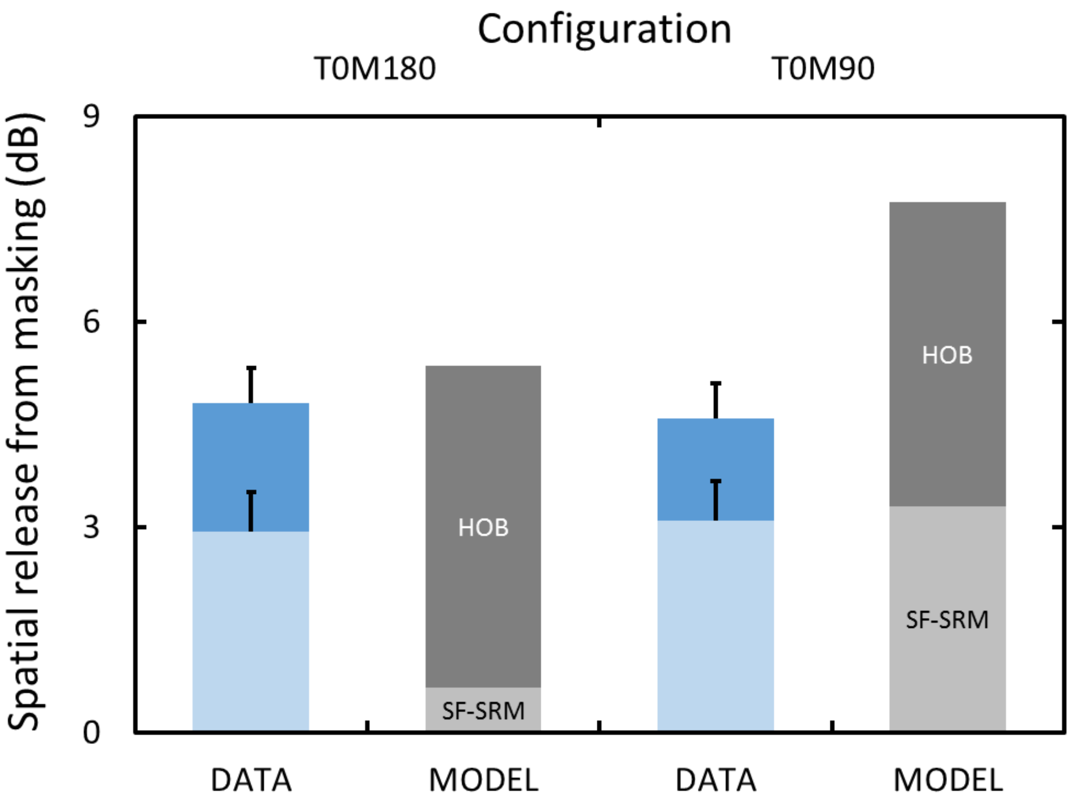


Figure 5.29: Bilateral CI users' audio-only SF-SRM (pale lower bars) and 30º HOB (dark upper bars) against model predictions. Error bars are standard error of means

The cumulated SF-SRM and HOB were 4.8 and 4.6 dB at T_0M_{180} and T_0M_{90} respectively, compared to the 5.4 and 7.8 dB model predictions. This is once again consistent with the suggestion that bilateral CI users deviated from facing the speech at H_0M_{180}. The larger deviation from model predictions at T_0M_{90} suggests that the model overestimates HOB for bilateral CI users. Indeed the discrepancy could not be accounted for by inaccuracies in head orientation. The following section may provide an explanation for this discrepancy.

5.4.2.5. Summation and squelch

Summation is defined as the advantage of hearing with two cochlear implants with identical signals arriving at the two sides. In other words, it is the improvement in H_0M_0 SRT, between situations with the better CI activated and with both CIs activated. Squelch is defined as the advantage activating the acoustically poorer CI for spatially separated sound sources. Squelch is traditionally measured in the H_0M_{90} configuration, where only the masker signal is subject to ILDs and ITDs. It is also measured here in the $H_{30}M_{180}$ configuration, where both speech and noise signals differ between ears.

Summation and squelch outcomes are plotted in Figure 5.30. Those were extracted from SRTs acquired with the audio-only protocol. An average summation of 2.9 dB (1 dB SE) was measured whilst squelch was 2.0 and 2.6 dB (0.5 and 1 dB SE) at H_0M_{90} and $H_{30}M_{180}$ respectively. Schleich et al. (2004) found average summation and H_0M_{90}-squelch of 2.1 and 0.9 dB across 19 users of the Medel Combi 40 or 40+ CI. Our measured benefits are larger, but are comparable within measurement error with Schleich et al.'s findings.

A within-subject T-test (2-tailed) comparing H_0M_0 SRTs with both CIs enabled to SRTs with the best CI enabled showed the summation effect to be significant ($t(7) = 2.84$, $p < 0.025$). The squelch effect was also found significant at H_0M_{90} ($t(6) = 4.05$, $p < 0.007$) and at $H_{30}M_{180}$ ($t(7) = 2.68$, $p < 0.032$).

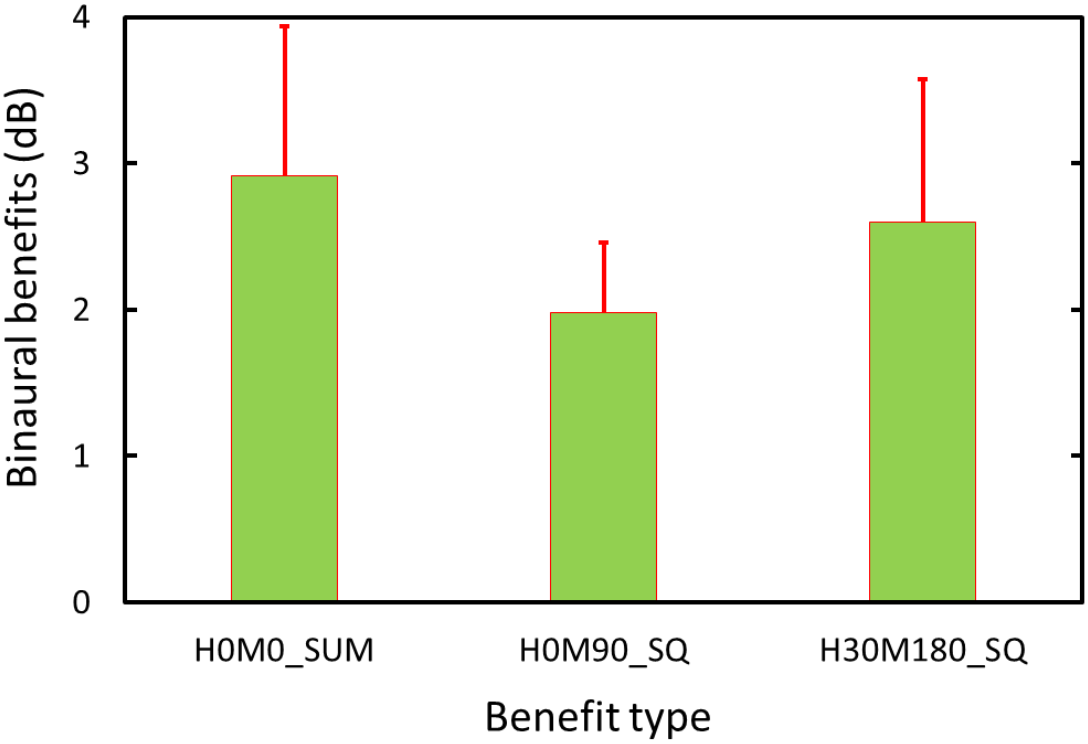


Figure 5.30: Bilateral CI users' summation (SUM) and squelch (SQ) measured in the H_0M_0, H_0M_{90} and $H_{30}M_{180}$ configurations. Error bars are standard error of means

5.4.3. Unilateral CI users

9 unilateral CI users first performed 3 or more repeat runs with the modified SRT protocol. All participants but one had omni-directional microphone settings. The data for that participant is presented later in a dedicated section. The remaining 8 participants' speech-facing SF-SRMs and additional 30° HOBs are plotted in Figure 5.31 for audio-alone and AV presentations against model predictions for the two spatially separated configurations. Since less than half of the participants performed the SRT task in the

additional, T_0M_{270} configuration, the outcome in this configuration will only be presented when run with the audio-only protocol.

5.4.3.1. Speech-facing SRM

As seen with other listener types, the T_0M_{180} SF-SRM was large (1.5 dB) compared to the BRIR-based prediction (0.5 dB). The difference between audio-only and AV SF-SRM is 0.4 dB, well within the standard errors (~1.1 dB). For T_0M_{90}, audio-only and AV SF-SRMs were 3.6 and 3.9 dB (0.9 dB SE) respectively. This compares with a model prediction of 3.5 dB.

An ANOVA for speech-facing SRT confirms that masker separation had a significant effect on SF-SRT ($F(2,14) = 9.90$, $p < 0.002$), with pairwise comparisons showing a significant H_0M_{90} SF-SRM ($p < 0.001$). However, due to large measurement errors, the H_0M_{180} SF-SRM did not reach significance ($p = 0.22$).

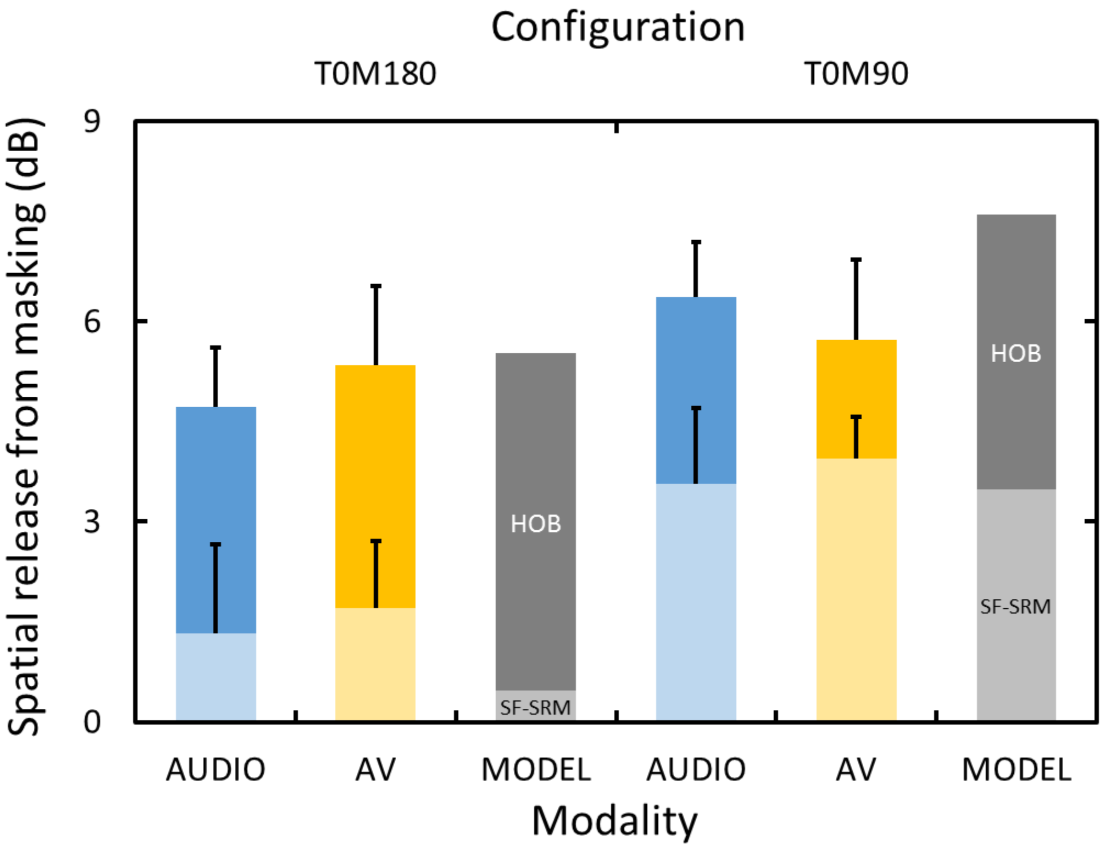


Figure 5.31: Unilateral CI users' SF-SRM (pale lower bars) and 30º HOB (dark upper bars) in audio and AV modalities against model predictions. Error bars are standard error of means

5.4.3.2. Additional 30º HOB

T_0M_{180} HOBs of 3.4 and 3.6 dB (0.9 and 1.2 dB SE) were gained in audio and AV, respectively. This was markedly lower than the 5.1 dB model prediction. T_0M_{90} HOBs of 2.8 and 1.8 dB (0.8 and 1.2 dB SE) were obtained in audio-only and AV respectively, compared to 4.1 dB predicted. Again, this was markedly lower than predicted. An ANOVA for SRM confirmed that head orientation had a significant, beneficial effect ($F(1,7) = 47.00$, $p < 0.002$) with no significant interaction with spatial configuration ($F(1,7) = 1.35$, $p = 0.28$).

The cumulated SF-SRM and HOB in the T_0M_{180} configuration were 4.7 and 5.4 dB in audio and AV respectively, compared to the 5.5 dB model prediction. The fact that cumulated benefits almost reached model predictions suggests that unilateral CI users, as other listener types, deviated from facing the speech at H_0M_{180}.

5.4.3.3. Lip-reading benefit

Figure 5.32 displays the bilateral CI users' LRB measured in each spatial configuration. The benefit ranged across participants and configurations from 0.3 to 11.0 dB, averaging 5.0 dB (0.8 dB SE on average). As found with unilateral users, this is significantly larger than the 2.8 dB found for the age-matched NH controls (see group comparison in Section 5.4.4.2). The LRB ranged from 4.2 to 5.5 dB across configurations and from 2.3 to 7.3 dB across participants.

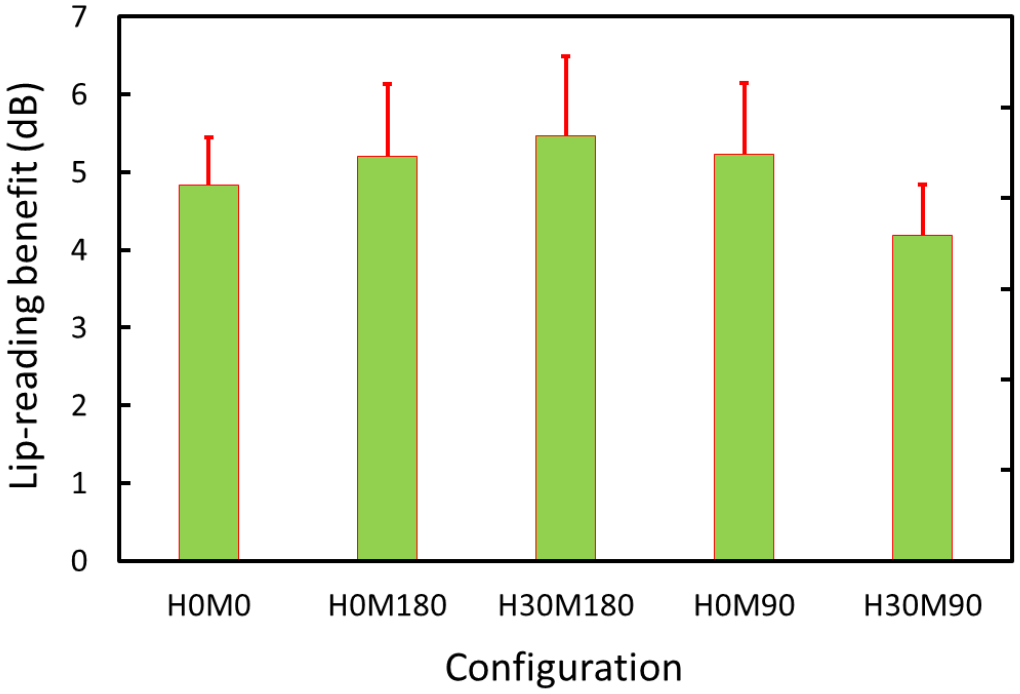


Figure 5.32: Unilateral CI users' LRB. Error bars are standard error of means

As for all other listeners, an ANOVA for SRTs confirmed the LRB to be significant ($F(1,7) = 80.43$, $p < 0.001$) with no interaction between modality and configuration ($F(4,28) = 0.29$, $p > 0.85$). LRB was also confirmed to be unaffected by a 30º head turn ($F(1,7) = 0.22$, $p = 0.65$).

5.4.3.4. Precise audio SRT measures

The 8 unilateral CI users with omni-directional microphone setting performed 6 repeats with the audio-only protocol. The speech-facing SF-SRM and additional 30º HOB are plotted in Figure 5.33. The additional T_0M_{270} configuration was chosen to show that, even when the masker is on the CI side, a significant 30º HOB could be obtained. The outcomes for the repeated conditions are consistent, within measurement error, with results obtained with the modified protocol.

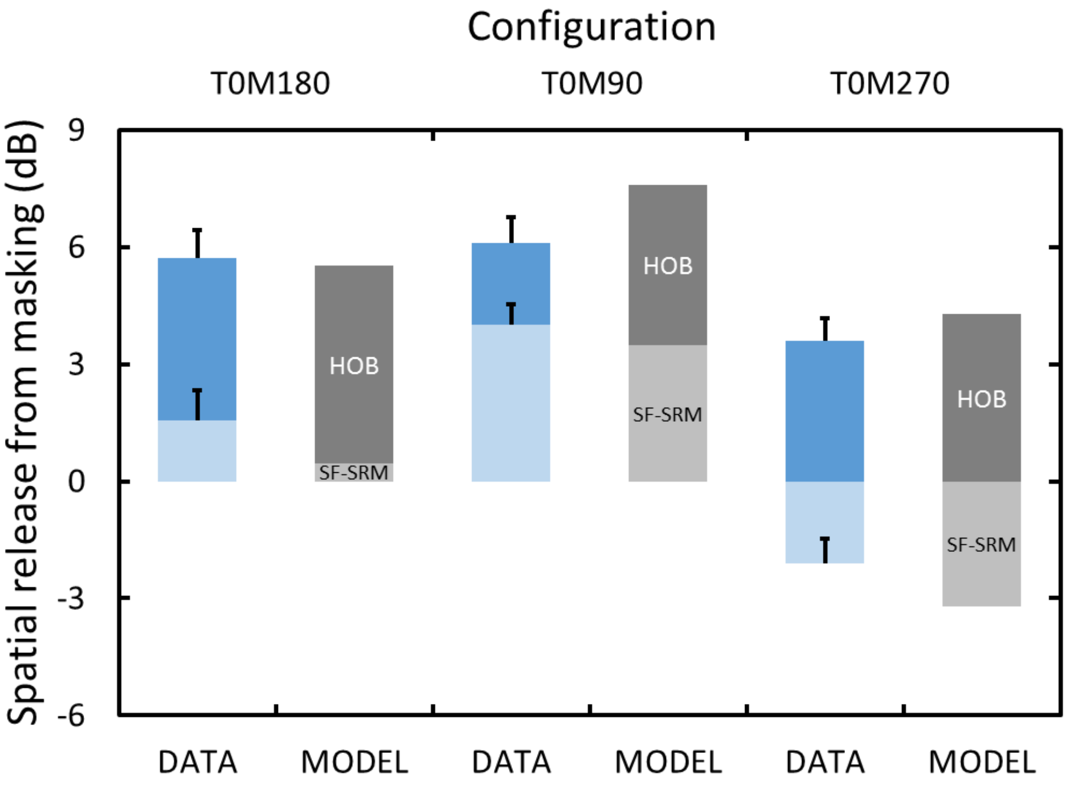


Figure 5.33: Unilateral CI users' audio-only SF-SRM (pale lower bars) and 30º HOB (dark upper bars) against model predictions. Error bars are standard error of means

SF-SRM was 1.6, 4.0 and -2.1 dB and 30º HOB was 4.2, 2.1 and 3.6 dB (0.64 dB average SE) at T_0M_{180}, T_0M_{90} and T_0M_{270} respectively. An ANOVA for speech-facing SRT confirmed that masker separation had a significant effect on SF-SRT ($F(3,21) = 30.55$, $p < 0.001$), with pairwise comparisons showing significant H_0M_{90} and H_0M_{270} SF-

SRMs (p < 0.012). The H_0M_{180} SF-SRM again did not reach significance (p = 0.084). An ANOVA for SRM confirmed a significant HOB (F(1,7) = 128.24, p < 0.001).

The cumulated SF-SRM and HOB were 5.6, 6.1 and 1.4 dB at T_0M_{180}, T_0M_{90} and T_0M_{270} respectively, compared to the 5.5, 7.6 and 1.1 dB model predictions. This suggests that the model provides good predictions for this population.

5.4.3.5. A directional microphone case

The 9th unilateral CI user had a Cochlear Ltd. ESPrit 3G processor. This unit has a built-in directional microphone, more sensitive to sound originating from the front of the user. We were interested here in investigating how directionality of sound pick-up would affect HOB. To that end, we considered the results for the audio-only protocol.

We were not able to acquire BRIRs in the test room with the ESPrit 3G directional microphone. The model predictions however needed to be adjusted to reflect the directional microphone effect. The model was first fed with two sets of HRIRs provided by Cochlear Ltd. One acquired with an omnidirectional microphone, the second with a directional microphone. The difference between the two HRIR-based predictions was then added to our BRIR prediction to form a new prediction. Of course this operation would not exactly result in BRIR-based predictions for a directional microphone in our test room, but that was the nearest approximation within our reach.

The SF-SRM and additional 30° HOB data are plotted in Figure 5.34 next to the directional-microphone-based predictions. When compared with unilateral CI users with omnidirectional microphones (Figure 5.33), it is immediately obvious that there is a very large 10 dB boosting by the directional microphone of the T_0M_{180} SF-SRM. Such a large change was expected with the masker directly behind the listener. With the masker at the side of the listener, at T_0M_{90} or T_0M_{270}, the change was much less marked (1.6 dB). HOB predictions were reduced by about 1 dB in all spatial configurations, which was reflected also in the HOB data.

SF-SRM was 12.0, 6.7 and -0.5 dB and 30° HOB was 2.2, 2.6 and 5.6 dB (1 dB mean SE) at T_0M_{180}, T_0M_{90} and T_0M_{270} respectively. An ANOVA for speech-facing SRT confirmed that masker separation had a significant effect on SF-SRT (F(3,12) = 35.44, p < 0.001), with pairwise comparisons showing significant H_0M_{180} and H_0M_{90} SF-SRMs (p < 0.008). The small H_0M_{270} SF-SRM was not found significant (p = 0.69). An ANOVA for SRM confirmed a significant HOB (F(1,4) = 36.24, p < 0.004).

The cumulated SF-SRM and HOB were 14.2, 9.3 and 5.1 dB at T_0M_{180}, T_0M_{90} and T_0M_{270} respectively, compared to the 11.8, 9.1 and 4.6 dB model predictions. This indicates that the model corrections led to good predictions for T_0M_{90} and T_0M_{270} but that the corrections applied were not sufficient for T_0M_{180}.

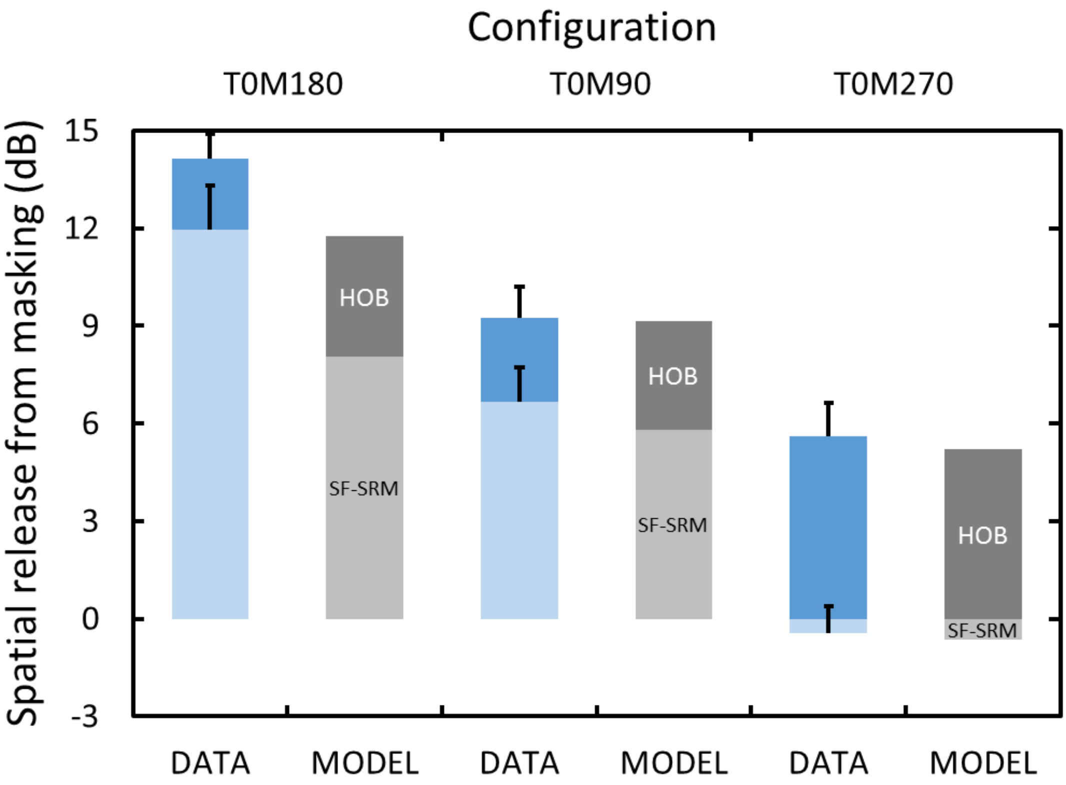


Figure 5.34: A unilateral ESPrit-3G CI user's audio-only SF-SRM (pale lower bars) and 30º HOB (dark upper bars) against adjusted model predictions. Error bars are standard error of means

5.4.4. Comparisons between listener groups and overall effects

5.4.4.1. Young adults versus age-matched normally hearing listeners

When comparing SF-SRM (across spatial configurations and modalities) between young adult and age-matched NH listeners, an ANOVA found no significant difference (F(1,18) = 0.35, p > 0.5). However, when comparing HOB between listener groups, significance was almost reached (F(1,18) = 4.07, p = 0.059). Only when considering the audio-only HOB did we find a significant difference in HOB between groups (F(1,18) = 5.54, p < 0.03), HOB going down with age from 4.46 to 3.27 dB, the drop being strongest (2.7 dB) in the T_0M_{90} configuration.

Comparing LRB between groups (across configurations), no significant age-related reduction in LRB was found (F(1,18) = 0.26, p > 0.6), as on average, young adults reached 3.0 dB against 2.8 dB for older NH adults.

5.4.4.2. CI users versus age-matched listeners

For SF-SRM and HOB comparisons, the lowest variability data from the audio-only protocol was used.

An ANOVA for SF-SRM across configurations did not show a significant difference between listener types ($F(2,23) = 0.97$, $p = 0.39$). However, it revealed a significant interaction between listener type and spatial configuration ($F(2,23) = 6.64$, $p < 0.005$). Indeed, whilst NH listeners and unilateral CI users consistently obtained, as predicted, significantly lower SF-SRM at T_0M_{180} (2.2 and 1.6 dB respectively) than at T_0M_{90} (5.1 and 4.0 dB respectively), bilateral CI users' SF-SRM differed little between configurations (2.9 and 3.1 dB respectively).

Comparing HOB between groups and across configurations revealed a significant listener group effect ($F(2,23) = 7.51$, $p < 0.003$). Pairwise comparisons again showed no significant difference between NH listeners and unilateral CI users ($p > 0.55$), but a significant difference between bilateral users and the other two listener groups ($p < 0.007$). Whilst NH listeners and unilateral CI users consistently obtained, as predicted, significantly larger HOB at T_0M_{180} (both 4.2 dB) than at T_0M_{90} (2.6 and 2.1 dB respectively), bilateral CI users' HOB differed little between configurations (1.9 and 1.5 dB respectively) and turned out significantly smaller than for the other two groups.

The SF-SRM and HOB model predictions did not differ greatly between unilateral and bilateral CI users, nor did T_0M_{180} SF-SRM or T_0M_{90} HOB predictions between listener groups. T_0M_{180} HOB and T_0M_{90} SF-SRM predictions for NH listeners, however, stood about 2 dB higher than CI-user predictions, due to the binaural unmasking component of the NH model of SRM. It therefore appears that age-matched NH listeners obtained less than half of their predicted binaural benefit, when compared to unilateral CI users. Furthermore, bilateral CI users stood out in that they did not follow the across-configuration trends in SF-SRM or HOB predicted by the model and seen with the other listeners. An explanation for why this may be the case will be offered in the Section 5.6 discussion.

An ANOVA for LRB across configurations showed a significant effect of listener type ($F(2,23) = 8.48$, $p < 0.002$). Pairwise comparisons showed that both CI user groups had a significantly larger LRB than age-matched NH listeners ($p < 0.01$) but that the difference between unilateral and bilateral CI users was not significant ($p > 0.3$).

5.5. Head-orientation advice given to UK CI users by professionals

Two short questionnaires were generated (see appendices 1-2), aimed at establishing what advice, if any, is currently given to CI users regarding head orientation with respect to an interlocutor in a noisy environment such as a restaurant. Multiple choice answers were made available regarding advice given/received (flagging if none was provided) and comment boxes gave respondents the option to specify alternatives to the proposed choices. The latter were randomised so as to avoid bias. Questions regarding the rationale behind any given advice were also asked. One questionnaire was designed for CI users, the other for the CI professionals we thought likely to provide communication advice to service users. The first was completed by 16 CI professionals (8 audiologists, 6 speech/language/hearing therapists and 2 teachers of the deaf). The second by 95 CI users (55 unilateral, 14 bilateral and 26 bimodal).

Of the 16 professionals, who all declared having given advice on this topic, 14 (88%) indicated their advice was frequently or always to face the speaker and never to turn the head away from them. 2 participants declared occasionally advising patients to turn their head away from the speaker.

A majority of the professionals who participated believed that facing the speaker is the best strategy for speech intelligibility. Factors that respondents selected as influencing their responses included, ‘ease of lip-reading’ (15), ‘microphone directionality’ (14), ‘ease of maintaining eye contact’ (11), ‘SNR at the better ear’ (10), ‘training, lectures/presentations or literature’ (7) and ‘social acceptability of orienting one’s head away from the speaker’ (3). From this small sample, it seems that professionals generally advise patients to face the speaker mostly because of preconceptions regarding lip-reading or microphone directionality.

Of the 95 CI users, 42 (44%) declared having never been advised on the subject at hand. Of those 53 who were given advice on it, all indicated they were at least occasionally advised to face the speaker, and 41 (77%) “frequently” or “always”. On the other hand 23 (43%) declared being advised at least occasionally to bring their best ear closer to the speaker and to carry on lip-reading, of which 16 (30%) “frequently” or “always”.

The CI users’ survey was broadly consistent with the professionals’ in that CI users declared being advised to face the speaker more often than to make use of head orientation. Moreover, when asked the reasons they recalled being given for the advice provided, those selected were ‘ease of lip-reading’ (43), ‘ease of maintaining eye contact’ (23), ‘benefit from their microphone directionality’ (22) and ‘SNR at the better ear’ (13). Again, this data points to professionals’ preconceptions on the requirements of lip-reading or directional microphones.

5.6. Discussion

Objective SF-SRM and HOB of unilateral and bilateral CI users and age-matched normally hearing listeners were measured as a function of masker separation and head orientation. SF-SRM and HOB were first measured with the ‘modified paradigm’ in audio and AV modalities: SF-SRM with the target directly ahead and a masker collocated with it (T_0M_0) or separated from it when placed behind (T_0M_{180}) or to the side (T_0M_{90}) of the listener; HOB with a modest 30º head orientation away from the target and in the separated spatial configurations. Audio-only SF-SRM and HOB were subsequently re-measured with higher precision. Both age-matched NH listeners and unilateral CI users broadly followed the model-predicted SF-SRM and HOB trends across configurations. However, bilateral CI users did not, as spatial configuration made little change to their SF-SRM or HOB. Regardless, all listeners were found to benefit significantly from a modest head orientation away from the speech. Age-matched NH listeners gained 2.6 to 4.2 dB and CI users 1.5 to 4.2 dB HOB. This HOB is particularly important to CI users, for whom noisy social settings present a severe challenge.

All listener groups had an inflated T_0M_{180} SF-SRM compared to model predictions. This was believed to be due to listeners having difficulty in exactly facing the target when required to do so. This meant that 1-2 dB of the measured T_0M_{180} SF-SRM may in fact have been HOB. Whilst this may account for a deflated measure of T_0M_{180} HOB in CI users or young NH listeners when compared to model predictions, it does not fully do so in age-matched NH listeners. This suggests that older NH adults suffer from a loss of binaural unmasking. This is consistent with recent reports of an age-related decline in the binaural processing of temporal envelope and fine structure (King et al. 2014; Moore et al. 2012; Hopkins & Moore 2011).

Bilateral CI users stood out in that their measured HOB was less than half of model predictions. At T_0M_{180} this may again be explained by inaccuracies in head orientations during testing. However, at T_0M_{90}, the HOB shortfall clearly requires another explanation since the overall SRM sits 3 dB lower than predicted. Additional measures of summation (2.9 dB at H_0M_0) and squelch (2.0 dB at H_0M_{90} and 2.6 dB at $H_{30}M_{180}$) from bilateral CI users were found to be somewhat larger than previously reported in the literature

(Schleich et al. 2004). Assuming bilateral CI users do not benefit from binaural unmasking, both summation and squelch are believed here to be due to the information provided to both CIs differing in spectral content, in a manner such that each CI spectrally complements the other to some extent. Indeed, bilateral CI users and particularly those having been implanted later in life, as was the case for all of our 8 participants, are unlikely to have equal nerve survival along their spiral ganglia. Moreover, it is common for some of the CI electrodes to be disabled, so as to prevent, for instance, unintended facial nerve excitation. Finally, insertion depth may vary between the ears. A plausible complementarity in spectral content across the ears would lead to a spectral summation effect as the second CI provides additional spectral information relevant to the speech. The same argument can be made with respect to squelch, the only difference being that in that case the acoustic signals at the ears already differ to some extent.

Since the model only includes better-ear listening for bilateral CI users, any spectral summation or squelch, by definition due to addition of a second ear, could cause a discrepancy between data and model predictions. The model indeed ignores the SNR at the poorer ear when it could still be relevant to speech intelligibility if it contains different or complementary spectral information at a sufficient SNR. We will now examine what may be happening to the SNR at the poorer ear and the impact this may have on bilateral CI users' SRM in the spatial configurations used in this study.

When the acoustic input to both ears is the same, summation improves SRTs. Since this is the case for both H_0M_0 and H_0M_{180}, one would expect spectral summation to improve SRTs equally in both configurations. The T_0M_{180} SF-SRM being computed by subtracting the H_0M_{180} SRT from the H_0M_0 SRT, summation should be cancelled out. The T_0M_{180} SF-SRM should therefore be unaffected by summation. A caveat has to be added to the above considerations, since any misalignment of the head would make H_0M_{180} a squelch case, in which case SF-SRM will be affected by summation not being exactly cancelled out in the SRT subtraction.

When the acoustic input to both ears differs, squelch can also improve SRTs, but only provided the poorer ear SNR is sufficient. Assuming a spectral squelch occurs, one would expect squelch at $H_{30}M_{180}$ to be somewhat smaller than summation at H_0M_{180} since the poorer ear's SNR is reduced by a 30º head turn. The $H_{30}M_{180}$ squelch measurement was however marginally smaller than the H_0M_0 summation measurement, which suggests that squelch had little to no effect on the T_0M_{180} 30º HOB. At H_0M_{90}, when compared to H_0M_0, although the better ear receives an SNR boosted by the head-shadow effect, the poorer ear that now faces the masker receives a much reduced SNR. This should reduce the spectral benefit of the poorer ear. The H_0M_{90} squelch was measured to be 1 dB lower than the H_0M_0 summation (be it not significantly), consistently with an SNR reduction at the poorer ear. Out of all the configurations examined, the loss of spectral squelch may be highest at $H_{30}M_{90}$ where the poorer ear receives the strongest masker and weakest target levels, thereby reaching the lowest SNR. It is plausible that in that worst (examined) case the spectral benefit of the poorer ear may be significantly eroded. Whilst a 30º head orientation was predicted to provide our bilateral CI participants with a 4.5 dB T_0M_{90} HOB, this benefit may have been offset by a loss of spectral squelch of up to 2-3 dB. We feel this is the most plausible explanation for bilateral users' T_0M_{90} HOB measuring as low as 1.5 dB. The fact that with an additional CI, bilateral users' $H_{30}M_{90}$ SRM measured 1.5 dB lower than unilateral users' further reinforces our interpretation of the data.

When comparing AV to audio-only SRM outcomes, the LRB for all listener groups was found to be independent of masker location. An additional 30º head orientation away from the speech had no effect on the LRB. CI users' LRB was measured to be 5 dB, 2 dB higher than age-matched listeners'. This suggests CI users have a greater facility for lip-reading, which is consistent with previous reports on the matter (Mitchell & Maslin 2007; Giraud et al. 2001; Rouger et al. 2007). Most relevant to this study is the fact that the lip-reading benefit that CI users rely on so much was not negatively impacted by a head orientation that was shown to provide a significant HOB. This is a win-win situation for CI users.

The case of a unilateral CI user that had a BTE unit with microphone directionality was also examined. Whilst SF-SRM was found to be boosted by up to 8 dB with the masker placed behind the listener, the HOB was significant and exceeded 2.2 dB in all spatial configurations. This illustrates that the (limited) directionality of directional microphones does not prevent their users from gaining a HOB. Although we could conclude this for directional microphones, we believe that it could also be the case when sound pick-up directionality is achieved through processing of two or more omnidirectional microphone inputs, as long as such directionality is not automatically focussing on the speech signal.

When attending to long audio-only and AV clips with diminishing SNR in the T_0M_0, T_0M_{180} and T_0M_{90} spatial configurations, a majority of CI users and age-matched NH listeners in an undirected paradigm did not make use of an effective head-orientation strategy. Only in 30, 25 and 10% of overall trials did age-matched NH listeners, bilateral and unilateral CI users respectively turn their heads. This was less than the 45% found in Chapter 4 with young NH adults in identical conditions. We had originally hypothesised that CI users would be more likely to make use of spontaneous head orientation and unilateral users more so than bilateral users. This is because we believed that the challenge that noisy environments present to CI users would lead to a stronger motivation to make use of all available benefits, including HOB. Instead, CI users were found to make less spontaneous use of head orientation than NH listeners (although not significantly so). Our survey of CI users and professionals may shed some light as to why, as it shows that UK professionals' preconceptions about lip-reading or directional-microphone requirements leads them to generally advise CI users to face their interlocutor and not deviate from that. Over half of our CI-using participants indicated having been given such advice. The LRB data from all our CI participants and the SRT data acquired from a directional-microphone CI user both indicate that professionals are not giving the best advice.

Instruction to explore head orientation unsurprisingly increased head turns whilst AV presentation reduced them, indicating that seeing the speaker had a suppressive effect on the listeners' tendency to orient their heads away from them. This may be due to listeners feeling the need to orient towards the focus of their attention whilst lip-reading, and/or may be a reflection of the speech-facing social norm affecting listeners' behaviours when the face of the speaker is visible. No categorisation of listening strategies was attempted. Subjective measures of SRM were found to increase significantly as a result of instruction. Both NH listeners and CI users gained larger HOBs, because instruction led them to make larger head turns away from the speech. Post-instruction subjective SRMs were still sub-optimal (see Figures 5.10 and 5.21), as larger head turns were either insufficient or incorrect. Subjective SRMs reached 4.3 dB on average for both listener groups. Objective SRMs measured with a 30° head turn showed that aged-matched NH listeners gained on average 2 dB more than CI users at that head orientation. This is consistent with the model predicting that about 2 dB more SRM is available to NH listeners (through binaural unmasking) at head orientations other than facing the speech. As both groups' subjective SRMs were equal, CI users appear to have reached more of the SRM available to them than NH listeners and in doing so, performed better than NH listeners in the post-instruction free-head task.

Considering that our CI-user participants improved their speech intelligibility within minutes of being given a simple instruction to explore the benefits of head orientation, the translational application of our findings to training designed to help CI users make the best of HOB is obvious.

5.7. Conclusion

Unilateral and bilateral CI users, as well as age-matched NH listeners were tested for the predicted head-orientation benefit they could reap from a modest head orientation away from a speaker with a single interfering noise placed directly behind them or to their side. Over and above the traditionally measured speech-facing spectral release from masking, adult unilateral and bilateral CI users (mean age 62) were found to significantly benefit from a 30º head orientation away from the speaker, with no detrimental effect on their superior lip-reading capacity. They were showed to obtain a SF-SRM and 30º HOB comparable to age-matched NH listeners, each in the 1.5 to 5 dB range, with their overall SRM reaching up to 6-7 dB. This demonstrates how CI users could boost their understanding of an interlocutor once they had chosen, for instance, a restaurant seat that placed the bulk of interfering noise behind them or to the side opposite to their better ear. Additional data from a CI user making use of a directional microphone showed that even in that case, the 30º HOB remained strong. In an undirected free-head paradigm, although listeners were spontaneously poor at making effective use of head orientation to aid their understanding of speech in noise, CI users and age-matched listeners significantly improved, immediately after instruction to explore the HOB available to them. As our bilateral-CI participants benefitted from a large summation/squelch effect that was reduced by head orientation, they performed best with the noise placed behind them. Our findings illustrate how a simple training program could provide CI users (and by extension other hearing impaired people such as hearing aid users) with a simple means to significantly reduce the severity of the challenge they experience when attending to speech in a noisy social setting.

6. General discussion

6.1. Summary of our research approach and findings

6.1.1. Research approach

This thesis grew from Culling et al. (2012). There, informed by the Jelfs et al. (2011) model of spatial release from masking (SRM), we demonstrated that CI users could benefit a lot more (20 dB) in speech-in-noise intelligibility from bilateral (over unilateral) implantation than had been typically reported in previous publications (e.g. Schleich et al. 2004; Laszig et al. 2004; Laske et al. 2009). Culling et al.'s predicted and measured SRM was the speech-intelligibility benefit of azimuthally separating a continuous speech-shaped noise (SSN) masker from a target. The benefit of adding the acoustically favoured ear to the acoustically penalised ear in a configuration where target and masker are azimuthally separated is typically referred to as the head-shadow effect (HS). In a target-in-front situation with a single masking noise at 90° (the T_0M_{90} configuration), previous studies reported SRM ranging 4-7 dB. In the symmetrically separated, $T_{45}M_{-45}$ configuration, just above 10 dB HS was measured (Laszig et al. 2004; Laske et al. 2009). The Jelfs et al. model predicted that the largest HS would be found in the $T_{60}M_{-60}$ configuration. CI users can listen with their better ear (BE) but exhibit negligible binaural unmasking (BU). Thus, the BE path of the model sufficed to predict which fixed-head situation would lead to the largest spatial release from masking (SRM) difference between those experienced by unilateral (UCI) and bilateral CI users (BCI). Prior to this thesis, all previous studies of SRM in CI users (but one, Van Hoesel 2015) were restricted to audio-only presentation and considered exclusively fixed-head situations. Since no published study had recorded spontaneous, free-head orientation in speech-in-noise listening situations for normally hearing (NH) listeners (let alone CI users), a natural starting point for this thesis was to perform a free-head, audio-only speech-in-noise NH listening experiment. We measured NH baseline maps of speech-facing SRM (SF-SRM) and head-orientation benefit (HOB), the SRM relative to SF-SRM. Optimum energetic masking (EM) was obtained by using SSN as the masker. Further developments introduced audio-visual (AV) presentations and the testing of young NH listeners, CI users and NH listeners age-matched to our CI participants.

6.1.2. Young NH audio SRM & listening behaviour

The pilot experiments of Chapter 3 link the speech-reception threshold (SRT) data used to generate SRM/HOB maps, to behavioural outcomes and subjective measures in a free-head task. An earlier approach had not been successful at motivating spontaneous listening head orientations with NH listeners (Brimijoin et al. 2012; pers. comm.). Mindful of there being a risk that none of the listeners might spontaneously turn their heads, we clearly separated in all the studies presented herein, the fixed-head SRT and free-head listening experiments. This distinction allowed us, in the free-head task, to give our listeners ample opportunity to think of, and opt for, a listening strategy that included head movements. In the free-head task, young NH adults attended to 4-minute speech clips in a gradually diminishing speech-to-noise ratio (SNR) until they flagged losing track of what was said. That point in each clip provided us with a subjective measure of SRT, together with the corresponding final head orientation. The fixed-head task followed an adaptive SRT measurement protocol and provided objective measures of SRT. The fixed- and free-head experiments were run in four matching spatial configurations (T_0M_{180}, T_0M_{150}, $T_0M_{112.5}$ and $T_0M_{97.5}$). These configurations were chosen on the basis of anechoic model predictions, hypothesising that listeners in the free-head task, should they choose to make use of head turns, would be sensitive to both the slope and the amplitude of SRM changes with head orientation. Spatial configurations were chosen to have a variety of HOB slopes, symmetrical or not about the speech-facing position. T_0M_{180} was chosen as the configuration predicted to provide the largest HOB, symmetrically so about the speech-facing orientation. The other three configurations provided asymmetrical HOB, positive one way, negative the other.

During the free-head tasks, young NH listeners were found to spontaneously turn their heads in 60% of trials. A few of those who moved leapt straight to the optimum orientation(s), as though they had already refined an effective strategy in their prior experience. Others gradually turned their heads, apparently scanning for intelligibility improvements. The remainder did not exhibit any seemingly structured or effective strategy. Overall, young (20 year-old) NH listeners were poor at spontaneously reaping the HOB available to them in an audio-only task. Those who repeated the task after instruction to make use of head orientation improved on average (some excelled), but some persisted in failing to gain a HOB or even moved to detrimental orientations. The listeners' behaviour was, against all expectations, apparently not dependent on the slope, nor on the symmetry of spatial configurations, nor did listeners exhibit a higher head rotation propensity when a high HOB was made available. Although a categorisation of head-orientation strategies was not strictly justifiable (we found no evidence that listeners' behaviours did not all form part of a behavioural continuum), we could conclude from the undirected and directed free-head experiments that training listeners to make use of head orientation could go some way towards helping them reach an optimum HOB in a noisy setting.

The SRT data and derived SRM/HOB maps further validated the Jelfs et al. model for HOB, once fed with binaural room impulse responses (BRIRs). Data and predictions were typically matched within less than 1.2 dB over a 12 dB SRM range. Only where SRM slope was maximum, and hence SRM was maximally sensitive to exact head orientation, were data points departing from predictions, but by at most 1.7 dB (overall RMS error of 0.77 dB across 32 conditions).

6.1.3. NH and CI audio and AV spontaneous listening behaviour

The experiments of Chapter 4 developed the paradigms of Chapter 3 by adding to the audio-only NH listeners' tasks equal numbers of AV presentations in the fixed- and free-head experiments. Most previous studies measured SRM or HS in the 'standard' H_0M_{90} (i.e. head-at-0º, speech-facing T_0M_{90}) and H_0M_{270} configurations. Although widely used, these configurations suffer from the *bright spot* effect, where the ear contralateral to the masker receives a locally high masker level due to the masker wave-front wrapping around the head and constructively interfering on the other side. This effect goes against the intuitive, but incorrect assumption that the ear contralateral to the masker would maximally benefit from the acoustic head shadow when the masker is at 90°. Moreover, the bright spot being highly localised, data acquired with the better ear placed on the bright spot will be very sensitive to the exact head orientation. Thus, such data will be prone to unusually high variability. Nevertheless, we felt that we had to link our data sets to the data from earlier studies by selecting T_0M_{90} as one of our base configurations. The T_0M_{180} configuration was retained from Chapter 3 for its high and symmetrical HOB slope and magnitude, whilst T_0M_{90} replaced the previous three asymmetrical configurations. Given that listeners were expected to make use of lip-reading during bimodal presentations, the SRT measurements were taken speech-facing as well as with a 30° head orientation that was predicted to provide the bulk (at $H_{30}M_{180}$) or the totality (at $H_{30}M_{90}$) of the available HOB, but we expected would not preclude lip-reading. The data acquired in the collocated (H_0M_0) configuration was used as a reference to compute the objective SRM data (from fixed-head SRTs) and the subjective SRM data (from the lowest SNR reached at final head orientation within each free-head track), whilst it served as a control for the behavioural data.

The Chapter 4 data served as a young NH baseline and the Chapter 5 experiments replicated the baseline experiments, but for three new listener groups: BCI users (8, age mean 54), UCI users (8, age mean 64) and 10 NH listeners age-matched to the CI users. Additional data was acquired for UCI users in the T_0M_{270} configuration, where the masker was placed ipsilateral to their CI, but despite a detrimental configuration, a 30° head orientation was still predicted to provide a substantial HOB.

As in the pilot experiments, the Chapter 4 experiments showed that young NH listeners were poor at spontaneously making use of an effective head-orientation strategy. The audio-only, spontaneous, free-head runs with young NH listeners repeated the earlier finding, whilst AV presentation significantly inhibited head movements. As a result, young NH listeners spontaneously moved their heads only 45% of the time. The aged-matched (mean age 59) NH listeners tended to move their heads even less (30% of the time) with a more marked AV inhibition. Against our expectation that CI users would make more spontaneous use of head orientation than NH listeners, BCI users spontaneously moved in only 25% of trials and UCI users only 10% of trials. AV inhibition had almost totally eradicated head movements for BCI and UCI users, confirming our expectation that more active and/or conscious lip-reading in CI users would reduce head movements.

As was found in the pilot study for young NH listeners, CI users and NH listeners (young and older) did not exhibit a clear set of head-orientation behaviours that could be categorised. A study with a much larger sample, would help establish whether behavioural categorisation is justifiable or whether behaviours belong to a continuum. Such a study could help tailor the design of training programs to the specific listeners' needs.

6.1.4. NH and CI objective and subjective measures of SRM

Figure 6.1 summarises the objective and subjective SRM data, placing both sets in the context of model predictions computed from BRIRs acquired in the test room. The top and bottom panels cover the T_0M_{180} and T_0M_{90} configurations respectively. The model predictions are displayed in the centre of each panel for the three listener types. To the left are the audio and AV objective measures, to the right the subjective ones. Each subset of data covers, for each of the four listener groups, the cumulative data from each

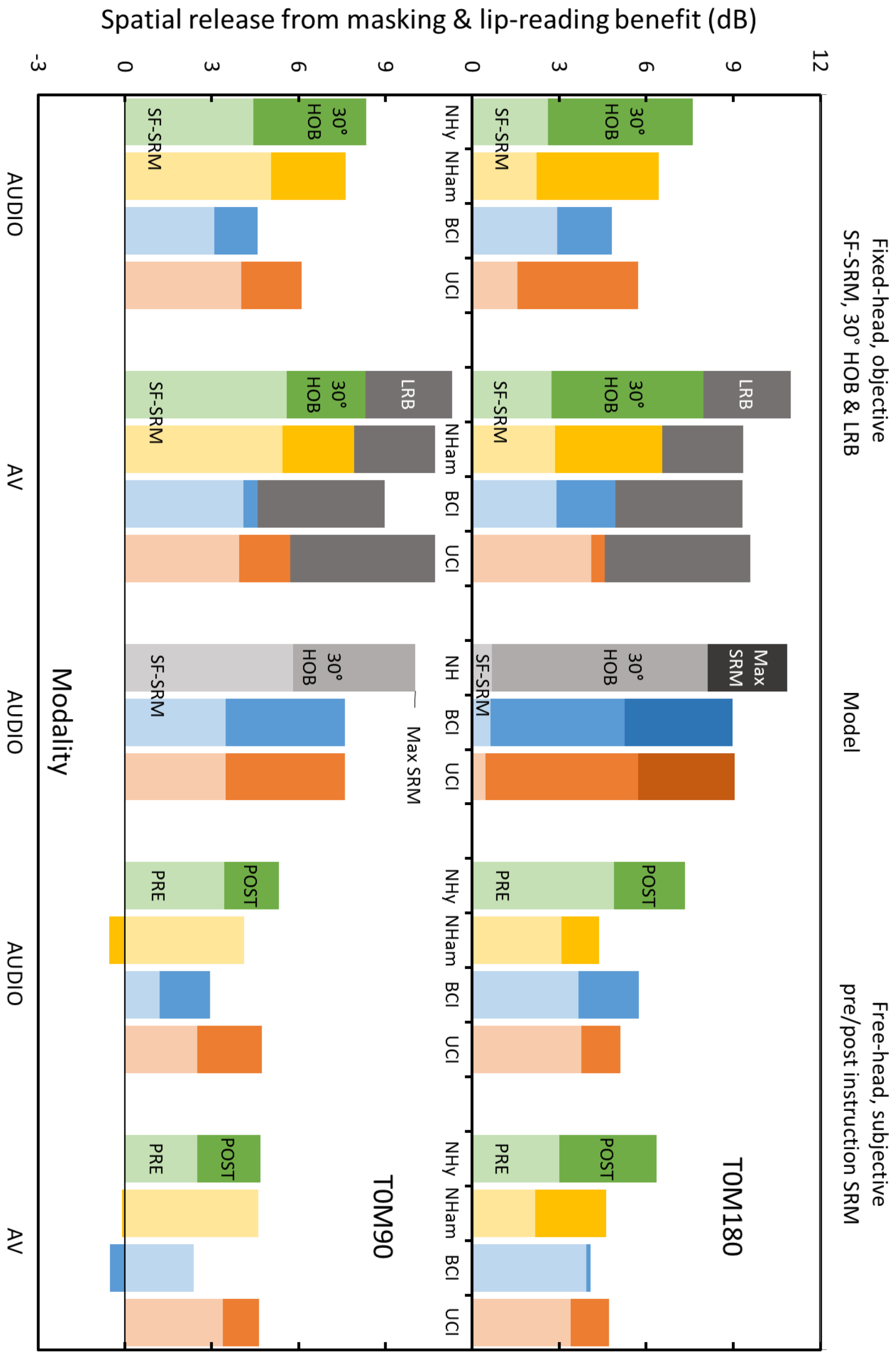


Figure 6.1: T$_0$M$_{180}$ (top) and T$_0$M$_{90}$ (bottom) measures of SRM and model predictions (centre). Objective measures (left) cumulate SF-SRM, HOB and LRB whilst subjective measures (right) cumulate pre-instruction SRM and post-instruction improvement

experimental set, i.e. SF-SRM and 30° HOB for the objective measures and pre-instruction SRM and post-instruction improvement for subjective measures. The LRB averaged over all conditions within a listener group (see following section) is also cumulated with the objective SRM data in the AV modality.

6.1.4.1. Objective SRM outcomes

At T_0M_{180}, the objective measures of SF-SRM were larger than predicted for all listener groups, probably because listeners were not quite facing the front. A benefit of about 1 dB was predicted to occur from only 5° of misalignment. At T_0M_{90}, NH listeners' objective SF-SRM data sat 0.5-1.5 dB below the 5.9 dB predicted. That of CI users, however, tended to exceed the predicted 3.5 dB (by up to 1 dB), perhaps an effect of the specific choice of BRIRs in the model computation. The objectively measured 30° HOB was lower than predicted for all listeners and in both spatial configurations. It however remained significant and substantial. At T_0M_{180}, the 3-4 dB shortfall in NH HOB may again have been due, in part, to the inaccuracy in head-orientation during SF-SRM measurements. At T_0M_{90}, the shortfall, averaged across presentation modalities, ranged between 0.9 dB (for young NH adults) and 3.5 dB (for BCI users). Age-matched NH adults typically exhibited the same SF-SRM as young adults, but their 30° HOB was reduced by 1-1.5 dB, consistently with a loss of BU with age (King et al. 2014; Moore et al. 2012; Hopkins & Moore 2011). This age-related loss of BU should however not affect BCI listeners if one assumes no BU is accessible to them. Some age-related loss of BE may also have been at play, which should affect age-matched NH and BCI listeners equally. However, the equal or larger-than-predicted SF-SRM measured across the board suggests that BE remains intact with age (providing normal hearing thresholds), for the most part.

6.1.4.2. Addressing the largest data-model discrepancy

The most obvious discrepancy between SRM model and data occurs at T_0M_{180}. Possible explanations include head asymmetry or misalignment or model under-prediction.

Considering NH listeners first, anechoic predictions of SRM versus head orientation (see Figure 2.6) reach a SRM minimum and a SRM-slope maximum (therefore a maximum SRM sensitivity to exact head orientation) when the listener faces the speech. The presence of moderate reverberation in our sound-deadened room would not change the trends. At T_0M_{180}, target and masker lie on the same cone of confusion (see Figure 1.1), which in this case coincides with the median plane. The presence of the pinnae favours somewhat the target faced by the listener, leading to less than 1 dB SRM. The manikin head used for HRIR and BRIR predictions is symmetrical and aims to represent an average human head. Any left/right asymmetry in real human heads will lead to a departure from the predicted minimum, similar to that due to a slight head orientation away from exactly facing the target. The average effect on SF-SRM of real pinnae shapes may not equate to that of the average pinnae. It could be substantially larger.

In the pilot study, the largest discrepancy between model and data was found where SRM was predicted to be most sensitive to head orientation, specifically when listeners faced the speech at T_0M_{180}, and at the corresponding SRM minima that were gradually shifted to increasingly negative head angles by moving the masker from the rear to the side of the listener (see Figure 3.13). Had SF-SRM increased due to random misalignment of the head, the variability in the data at these points would have been maximum. However, it was not the case for all masker separations. Instead, the data variability suggested a systematic error in head orientation, in other words a head orientation accuracy issue, which would lead to increased T_0M_{180} SF-SRM without inflation of variability. Other than the model failing to predict an unknown summation effect at T_0M_{180}, some sort of attentional effect that favours the target speech in front over a masker coming from the rear or a combination of both, misalignment of the head orientation during the SRT runs is the most likely explanation for the model's under-prediction. Any tendency listeners may have had within a run to let their heads drift away from facing the speech would have led to a lowering of the thresholds that the adaptive tracks converged on, without necessarily increasing run-to-run threshold variability.

With respect to model validation and the resulting trust in the accuracy of the model, the SRTs measured in the pilot study (Figure 3.13) led to a best SRM model fit (0.77 dB RMS error) with most of the data points matching predictions within less than 1.2 dB.

For CI users, additional sources of T_0M_{180} SF-SRM discrepancy between data and predictions could be considered. The microphone position on the BTE processor should slightly favour sounds coming from the rear. This is best illustrated by the backward shift in the butterfly SRM pattern with head orientation at T_0M_{180} (see Figure 2.6) due to the microphone position on a Siemens Acuris BTE unit. However, such a shift is symmetrical about the median plane (in binaural, symmetric hearing) and has no effect on SF-SRM predictions (considering either implantation side or both combined). Some of the CIs used by our participants may have some built-in directionality that cannot be controlled by the user, even when microphone directionality is not selected in programmable settings.

Cochlear, for instance, fitted at some point moderately directional microphones to the BTEs (Freedom processor) and later might have maintained some directionality in their programs, even when directionality is not selected. However, users of Cochlear BTEs did not exhibit significantly larger T_0M_{180} SF-SRM than users of equipment from other manufacturers. Therefore, microphone directionality is unlikely to be a source of SF-SRM discrepancy.

If the model was at fault, given that the effect is common to all listener groups, the BU path should be excluded since BU is not accessible to CI users. The BE contribution to SRM, purely acoustic by definition, cannot explain the T_0M_{180} SF-SRM discrepancy observed. The summation and squelch measured in our bilateral CI participants was larger than that reported elsewhere and was offered as a potential explanation for their HOBs being smaller than that of unilateral CI users. We cannot exclude the possibility that summation may be larger at T_0M_{180} (where it was not measured) than at T_0M_0, although we fail to imagine what mechanism could give rise to a difference. Moreover, summation could not possibly contribute to the data from unilateral CI users.

Overall, a systematic error in head orientation, where SRM is most sensitive to exact head orientation, perhaps coupled with a slight forward body or neck movement and downward head orientation (which could also increase SF-SRM), remains the most likely reason for T_0M_{180} SF-SRM being larger than predicted for all listeners. Re-testing with an orientable chin-rest to maintain correct head orientation during testing would help confirm our suspicions.

6.1.4.3. Subjective SRM outcomes

When viewed in the context of both the predicted and measured SRM, pre-instruction subjective measures were consistently small in all groups. All groups were far from having spontaneously reached their full HOB potential. Naturally, instruction to make use of head orientation significantly increased head movements for all groups. As a result, subjective SRM improved for all groups overall. At T_0M_{180} and in the audio-only modality, this was to be expected for the binaural listeners since an increase in the magnitude of head-turn either way was predicted to augment HOB (up to 60°, see Figure 5.3). In the same condition, UCI users' SRM was predicted to monotonically span -9 dB to +9dB over -60° to +60° of head orientation, head turns being beneficial one way and detrimental the other. UCI users presumably benefitted from an unequivocal HOB trend, and a perhaps simpler task than binaural listeners' since they made a judgment on head-orientation effectiveness from the SNR at a single ear. Their post-instruction SRM was only 0.6 dB shy of their 30° objective SRM (5.3 dB). Young NH listeners and BCI users also improved significantly post-instruction since they more or less reached their 30° objective SRM (7.6 and 5.7 dB, respectively). Older adults, however, were comparably far from reaching that performance (at 4.4 dB vs. 6.4 dB 30° objective SRM). Not only had they less HOB available to them than their younger counterparts, they seemed to experience more difficulty in reaping it.

At T_0M_{90}, young NH adults improved post-instruction, although they were still 3-3.5 dB away from obtaining their optimum SRM (8.3 dB measured at 30° head orientation). UCI users performed better in comparison as they were at most 1 dB short of their (30°) optimum SRM (5.9 dB). HOB was also reduced for older NH listeners and BCI users by partial and total loss of BU, respectively, when compared to young NH listeners. The resulting reduction in SRM slope and amplitude at T_0M_{90} might partially explain the tendency exhibited by these listeners, post-instruction, to not make effective use of head turns. Indeed, they often worsened their subjective SRM by turning their heads the wrong way, seemingly thinking that turning away from the noise direction would help, when it actually made matters worse. Their judgment may have been further hindered by the fact that turning the wrong way would not lead to large changes in SRM (see Figure 5.3 in Chapter 5).

6.1.5. Reliance on lip-reading and impact on subjective SRM

The objective data shows that lip-reading improves SRTs for all groups. On average, the lip-reading benefit (LRB) of NH listeners was 3 dB and that of CI users 5 dB. Crucial to these studies was the fact that the LRB was independent of head orientation for all groups, with the head moved from 0º to 30º. This confirmed that a 30º HOB could be obtained in a manner that left the LRB intact. Thus, a sidelong look sufficed to maintain the LRB at a normal level (assuming that facing the speaker is the social norm).

Figure 6.2 displays CI users' and age-matched NH listeners' individual LRB data as a function of their H_0M_0 audio SRTs. A linear regression analysis of LRBs versus H_0M_0 SRTs showed a negative correlation between LRB and proficiency ($r = 0.66$, $t = 4.31$, $p < 0.001$). This is not surprising since an elevation in their SRTs will motivate an individual to improve their lip-reading skills. Every 6 dB in SRT elevation was partially compensated for by 1 dB improvement in LRB. Since talkers differ in the ease with which they can be lip-read, the regression slope of data acquired with a different talker could be significantly different to the slope we found. One might expect that the easier it is to lip-read the talker, the higher the slope (AV SRTs could even diverge to infinitely low when an expert lip-reader understands the speech from a familiar talker without any auditory input). Thus, for easier-to-lip-read talkers (e.g. more familiar interlocutors), LRB might go much further towards compensating for the elevation in SRT CI users suffer from. Previous studies also showed that LRB is highly dependent on the ease of lip-reading of the sentence material (Macleod & Summerfield 1987). To date, it appears that nobody has established whether material and talker contributions to the ease of lip-reading are independent or interact.

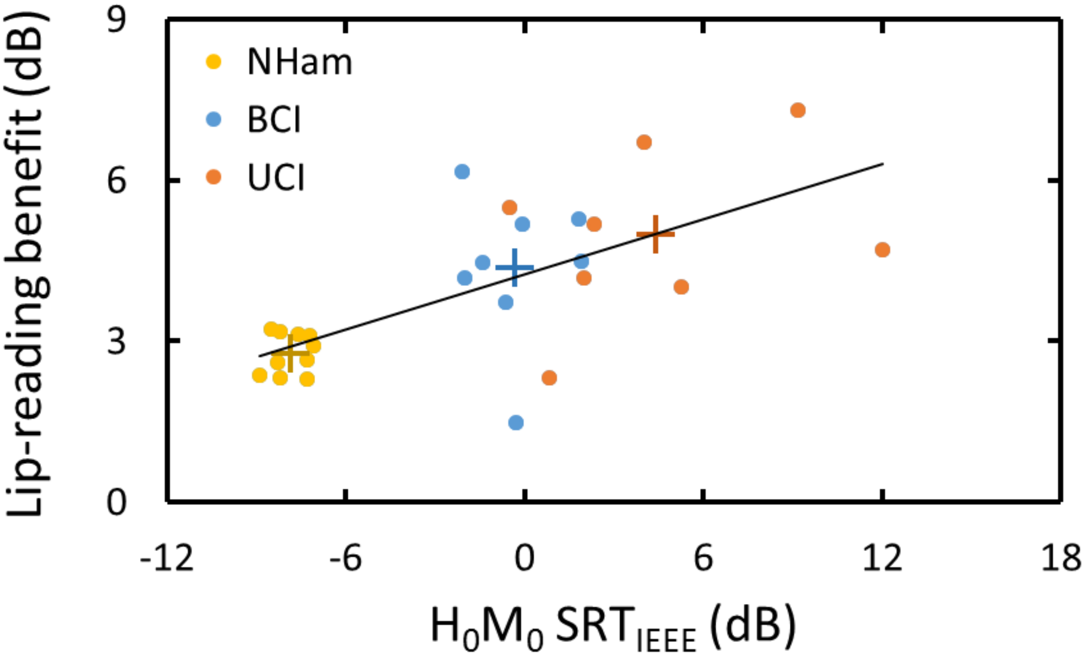


Figure 6.2: Lip-reading benefit versus speech-in noise understanding proficiency for bilateral and unilateral CI users and age-matched NH listeners

Although the subjective data in Figure 6.1 shows lower SRM means in the AV than in the audio-only condition, the reader should bear in mind that we defined AV SRM as the SRT improvement from the collocated situation *in the AV modality*. The LRB may have been reduced or lost for some participants by excessive head orientation away from the speech, thereby reducing the apparent AV SRM.

6.1.6. Relevance of microphone directionality

The special case of a UCI user that used a directional microphone setting demonstrated how, by suppressing sound waves coming from the rear, the T_0M_{180} objective SF-SRM was boosted by as much as 10.5 dB. The T_0M_{90} and T_0M_{270} SF-SRM values were increased by only 1.5 dB. In other words, if the masker were placed in the frontal hemifield, SRM was hardly affected by what is in reality a very limited microphone directionality. Microphone directionality is indeed not at all the ‘beam’ or ‘focus’ that manufacturers’ marketing material make it out to be. Just as importantly, and as a result of the limitations inherent to the typical cardioid or hyper-cardioid patterns of directionality, a significant 30° HOB remained in all three configurations. The T_0M_{180} HOB was halved, but the T_0M_{90} HOB was unaffected and the T_0M_{270} HOB was almost doubled. A bespoke study of the effect of different types of microphone directionality on HOB would be welcome.

6.2. Importance of our findings to the hearing impaired

CI users are known to struggle to understand speech in noisy social settings. Despite the restoration by CIs of the interaural level difference (ILD) cue and despite all the recent efforts made to restore access to interaural time delays (ITDs) at low frequencies, CI users exhibit negligible BU and pitch cues are limited by the relatively sparse encoding of sound by CIs. As a result, CI users’ only benefit from HS and LRB effects, both BU and F0 differences being inaccessible. Dip-listening is also much harder for CI users. Given the limited cues available to CI users, any benefit they can gain from guidance about how to make the best use of HS and LRB is extremely important to them. Such guidance can easily make the difference between their being socially isolated and their being able to actively enjoy social interactions. As this is true for a familiar, easier-to-lip-read interlocutor, it is even more critically important for unfamiliar, harder-to-lip-read interlocutors. Whilst the research presented herein focussed on CI users, it can equally well serve to help other HI listeners, whether partially and/or unilaterally deaf. Since BU represents a small part of a NH listener’s SRM and HI listeners often exhibit a reduction in BU, the conclusions drawn from the present studies can be directly applied to HA users as well as unaided HI listeners.

Chapter 5 presented the results of two UK surveys that revealed how, in the majority of cases, CI users are advised by CI professionals to face the speaker. Establishing why can serve to dispel erroneous beliefs. Assumptions that facing the speech is essential to lip-reading or critical to the optimum use of microphone directionality seemed to be the main culprits. More important is to think of how evidence-based guidance and training programs could stem from our and others’ research. The objective data presented in Chapter 5 proves that, given a favourable acoustic situation and a conducive social setting, CI users could first optimally position themselves (e.g. chose the best restaurant or lecture hall seat), then make use of head orientation to maximise their SRM. The subjective data from the free-head experiments demonstrates that, with a simple instruction to explore the benefits of head orientation, CI users gained significant extra HOB, and do so within minutes of a first, undirected trial. The objective data shows that significantly more HOB can be obtained with a refined strategy, and so without any loss of LRB. Specific guidance regarding the assessment of an acoustic scene and the optimum combination of head orientation and lip-reading could go much further. Simple training can easily be envisaged that does not require expensive equipment (our mobile lab cost less than £400, excluding the computer). Most ENT centres include an audiology suite that would lend itself to such training. The translational path is obvious, but our fundamental studies need to be complemented with additional ones that demonstrate that SRM can be obtained in more challenging situations. Seeing the speaker is more important to CI users than to NH listeners, not only for lip-reading, but also to reinforce and complement the auditory information from the target voice. Visual cues help CI users discriminate between a target and competing voices. In that sense, audio-only studies have a limited relevance to real-life situations and the multiplication of free-head, audio-visual studies of SRM would be very welcome.

6.3. Foreseen extensions to this thesis

Since no prior studies specifically measured CI users' HOB or their free-head, spontaneous listening behaviour and performance, a first investigation had to be designed with a restricted number of parameters. Those purposely excluded from the studies herein were reverberation, fluctuating maskers (including interfering voices), multiple and/or diffuse maskers and distance to the target speaker. Natural extensions to this thesis would consist of introducing the above parameters, one by one and in combination. One extension would explore the impact of reverberation on SRM/HOB, first with a single SSN masker. Another would investigate the impact of making use of a single, spatially-separated voice interferer, with forward and reverse speech, and the same or a different voice from the target, so as to separate informational masking from the exploitation of dip-listening or F0 differences (however limited for CI users). Another extension would investigate the effect of the spatial distribution of multiple SSN maskers. In gradually building up the auditory scene complexity, one would eventually simulate a cocktail party situation, including reverberation and multiple, spatially distributed voice interferers. Such a gradual build-up would help decipher which factors independently affect SRM (and HOB) and which interact. Van Hoesel (2015) simulated a cocktail party for CI users with audio and AV presentations. The author's aim was to demonstrate much larger benefit of bilateral (over unilateral) implantation in a dynamic, multi-talker situation that included the reverberation of a cafeteria. Bilateral listeners' SRTs were 5 to 15 dB lower than unilateral listeners' and the LRB was measured at typically 5 dB. As per Hirsh's (1950) original free-head SRM investigation, head orientations were not measured, nor could the effect of the different cues be unambiguously separated. The already extensive investigation of the cocktail party problem in NH listeners, HI listeners and CI users informs us on how CI users can benefit from the auditory and visual cues available to them, but the effect of spatially separated voice interferers on the intelligibility of a visible speaker by a CI user in a reverberant environment has not yet been fully broken down into its principal components. Models of SRM such as the Jelfs et al. (2011) model can provide essential guidance in reaching the most compelling evidence, as was demonstrated here and in Culling et al. (2012). The Jelfs et al. model, coupled with the preliminary studies presented herein, also lay a clear path towards designing efficient training programs, aimed at helping CI users make the best of a given acoustic setting. Conversely, the model and its computational efficiency lend themselves to helping architects and interior designers create spaces that optimise the acoustic comfort of NH listeners and the acoustic accessibility to HI listeners of social settings such as a restaurant. Although the model does not predict the potential for informational masking or exploitation of F0 differences with interfering voices, it provides the key guidance in terms of positioning and head orientation, guidance that would not be changed if these additional effects were present.

6.4. Reverberation and multiple interferers

We saw in Chapter 3 how the moderate reverberation of our sound-deadened testing rooms reduced the maximum HOB attainable by CI users from a predicted 13.5 (anechoic) to 9 dB (measured). In Culling et al. (2012), we explored the predicted effect of reverberation on HS. The most reverberant environments where we measured impulse responses (a kitchen and a stairwell) were predicted to broadly halve the HS predicted for our Cardiff testing room, whether in the H_0M_{90} or in the $H_{60}M_{120}$ configuration. Thus, in first approximation, one can expect the worst reverberant environments to reduce SF-SRM and HOB significantly. That said, SRM degrades surprisingly gracefully with reverberation as the pick-up effect of the head provides a baseline benefit that should always improve the received level of a nearby voice. A reduction in SRM would hinder CI users' ability to make use of an effective positioning and head orientation strategy, not only because HOB slopes would be proportionally reduced, but also because the assessment of the auditory scene necessary for optimum positioning in an unknown environment would be more challenging. The effect would be worst for unilateral CI users, since their sound source localisation is much less accurate. Luckily, Culling et al. showed that the HS predictions for two cafeterias remained close to those obtained for the sound-deadened room, and although predictions for two living rooms and a teaching room showed a significant SRM reduction, one would hope that CI users have more opportunity in such environments to control noise and voice interference.

In the multiple SSN interferer case, understanding the target could become very challenging for CI users. Again here, to make the situation easier they need to optimally position their better ear for maximum SNR. Positioning themselves in the room in such a way that the bulk of the interfering sources are placed in the rear hemifield (as they face the talker) will allow them to exploit their HOB. Only when the talker is positioned in the middle of interfering sources is there little hope of exploiting SRM. The only option CI users have is to ask the talker to change position.

Hindrance caused by a single interfering talker can, as for a single SSN interferer, be minimised by a CI user positioning themselves between the target and the interferer. The use of directional microphone(s) or processor setting will go a long way towards helping the situation, without impacting HOB too much. Multiple interfering talkers spread out in a room seriously complicate the situation. Indeed, informational masking can take place that may be particularly difficult for CI users to escape, since they do not have the ability to exploit F0 differences, nor can they combine ITD and F0 cues. Again, their only hope is to ask their interlocutor to move to a corner of the room, so the bulk of interferers are as far away as possible and kept in their rear hemifield.

Increasing the target speaker distance, when reverberation is present, will temporally blur the target voice, in addition to making lip-reading less accurate. Both will reduce speech intelligibility, regardless of the presence or absence of maskers. The Jelfs et al. model assumes that the talker is close by. In that sense, it applies itself well to HI listeners, who will always try to get close to the speaker, if only to have a better chance at lip-reading them.

Multiple interfering talkers in a highly reverberant room is clearly the worst situation for CI users (for all HI listeners), because a high wall reflectivity prevents them from getting away from competing voices reflected from all directions. Short of finding a less reverberant corner in that room, CI users' only hope is for architects and interior architects to design spaces with hearing accessibility in mind.

6.5. Concluding remarks

Despite the relatively artificial context of a testing room, of the presentation of speech over loudspeakers and of the presentation of the speaker's face on a video monitor, our CI participants showed great interest in our research. This was not just because the benefit they could reap from applying what they learned during and around our testing sessions was immediately available to them, but also because experiencing the HOB dispelled assumptions that they or CI-clinic professionals around them held. This helped them feel freer to explore how much better they could interact with interlocutors. A few of our participants volunteered feedback on how they got on when practicing in real-life environments. An excerpt from one of the testimonies we received sums it up:

> *" ...thank you for the advice about where and how to sit in noisy situations. ...in a pizzeria in Cannes, ... I chose to sit with my back to the other diners, facing Peter's colleague and Peter. ... at about 20 deg Russell's voice started to become louder and clearer, this continued to about 30 deg. It was as if someone was turning the volume wheel on my processor... "*

Incidentally, the above testimony came from the UCI user who used a directional microphone setting.

In addition to BCI users typically exhibiting lower SRTs in noise than UCI users, they also benefit from much more flexibility in terms of positioning in a room. Moreover, they seem to cope better with dynamic, multiple interfering talker situations because they are quicker at matching the visual, articulation cues with the auditory speech (Van Hoesel, 2015), presumably so because they can localise the source of a voice more easily or faster than UCI users can. However, some of our BCI participants exhibited larger-than-previously-reported summation and squelch, the extra benefit of adding a second ear to the better ear when speech and noise are collocated or separated, respectively. We expected this may be the case in more mature BCI users (those more likely to have dead regions along their spiral ganglia), since any spectral asymmetry/complementarity between the ears would improve speech intelligibility. A potential drawback of spectral squelch could be a reduction in HOB (see Chapter 5).

With or without spectral squelch, the HOB trend with head orientation was predicted to be less marked for BCI users than for UCI users (in all separated configurations but T_0M_{180}, see Fig 5.3). Thus, BCI users may find exploiting head orientation a more difficult task than UCI users. The complication for BCI users of having to make a judgement on optimum head orientation based on auditory input to both ears was clear. Indeed, more than half of the time, BCI users turned their heads away from the noise post-instruction at T_0M_{90}, thereby reducing their speech intelligibility. Instead, pointing their heads between the noise and the target directions was the only way to improve their speech intelligibility. They turned away from the noise, as though to reduce the noise level in the ear ipsilateral to the noise, rather than concentrate on the ear providing the best SNR. UCI users not only benefitted from a clearer HOB trend, their judgment of optimum head orientation was based on reaching the best SNR at their *only* ear. Their task being simpler may partly explain why they performed best post-instruction. Their superior post-instruction performance may also be explained by the fact that unilateral listeners in everyday life need to make more use of head orientation in order to appreciate an auditory scene, localisation being practically impossible for them without head movements. The above considerations suggest that, for any training to have optimal outcomes, CI and HA users, other HI listeners, and particularly binaural listeners, should be introduced to the concepts of head shadow and better-ear listening prior to any training sessions. Early provision of such information would ensure that patients fully appreciate that the key to their success is to maximise the SNR at the ear acoustically-favoured by their head orientation, not just to maximise the target level in one ear or minimise the noise level the other.

References

American National Standards Institute, 1997. American National Standard: Methods for Calculation of the Speech Intelligibility Index. *S3.5-1997 (R2012).*

American National Standards Institute, 1969. American National Standard: Methods for the Calculation and Use of the Articulation Index. *S3.5-1969 (R1986).*

Aronoff, J. et al., 2011. The Effect of Different Cochlear Implant Microphones on Acoustic Hearing Individuals ' Binaural Benefits for Speech Perception in Noise. *Ear and Hearing*, 32(4), pp.468–484.

Aronoff, J. et al., 2010. The use of interaural time and level difference cues by bilateral cochlear implant users. *The Journal of the Acoustical Society of America*, 127(3), pp.EL87–92.

Baer, T., Moore, B. & Kluk, K., 2002. Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies. *The Journal of the Acoustical Society of America*, 112(3), pp.1133–1144.

Bench, J., Kowal, Å. & Bamford, J., 1979. The Bkb (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children. *British Journal of Audiology*, 13(3), pp.108–112.

Beranek, L., 1947. The design of speech communication systems. *Proceedings of the IRE*, 19, pp.435–441.

Bergeson, T., Pisoni, D. & Davis, R., 2005. Development of Audiovisual Comprehension Skills in Prelingually Deaf Children With Cochlear Implants. *Ear and Hearing*, 26(2), pp.149–164.

Bernstein, L., Auer, E. & Takayanagi, S., 2004. Auditory speech detection in noise enhanced by lipreading. *Speech Communication*, 44(1-4), pp.5–18.

Beutelmann, R. & Brand, T., 2006. Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 120(1), pp.331–342.

Beutelmann, R., Brand, T. & Kollmeier, B., 2009. Prediction of binaural speech intelligibility with frequency-dependent interaural phase differences. *The Journal of the Acoustical Society of America*, 126(3), pp.1359–68.

Beutelmann, R., Brand, T. & Kollmeier, B., 2010. Revision, extension, and evaluation of a binaural speech intelligibility model. *The Journal of the Acoustical Society of America*, 127(4), pp.2479–97.

Bilger, R. & Hirsh, I., 1956. Masking of Tones by Bands of Noise. *The Journal of the Acoustical Society of America*, 28(4), pp.623–630.

Boothroyd, A. & Nittrouer, S., 1988. Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of America*, 84(1), pp.101–114.

Bradley, J. & Bistafa, S., 2002. Relating speech intelligibility to useful-to-detrimental sound ratios (L). *The Journal of the Acoustical Society of America*, 112(1), pp.27–29.

Brand, T. & Kollmeier, B., 2002. Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America*, 111(6), pp.2801–2810.

Breebaart, J., van de Par, S. & Kohlrausch, A., 2001a. Binaural processing model based on contralateral inhibition. I. Model structure. *The Journal of the Acoustical Society of America*, 110(2), pp.1074–1088.

Breebaart, J., van de Par, S. & Kohlrausch, A., 2001b. Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters. *The Journal of the Acoustical Society of America*, 110(2), pp.1089–1104.

Breebaart, J., van de Par, S. & Kohlrausch, A., 2001c. Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters. *The Journal of the Acoustical Society of America*, 110(2), pp.1105–1117.

Bregman, A., 1993. Auditory scene analysis: Hearing in complex environments. In *Thinking in sound: The cognitive psychology of human audition*. pp. 10–36.

Bregman, A., Liao, C. & Levitan, R., 1990. Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian journal of psychology*, 44(3), pp.400–413.

Brimijoin, W., McShefferty, D. & Akeroyd, M., 2010. Auditory and visual orienting responses in listeners with and without hearing-impairment. *The Journal of the Acoustical Society of America*, 127(6), pp.3678–3688.

Brimijoin, W., Mcshefferty, D. & Akeroyd, M., 2012. Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task. *Hearing research*, 283(1-2), pp.162–168.

Brokx, J. & Nooteboom, S., 1982. Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, 10, pp.23–36.

Bronkhorst, A., 2000. The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica*, 86, pp.117–128.

Bronkhorst, A. & Plomp, R., 1990. A clinical test for the assessment of binaural speech perception in noise. *International Journal of Audiology*, 29, pp.275–285.

Bronkhorst, A. & Plomp, R., 1992. Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *The Journal of the Acoustical Society of America*, 92(6), pp.3132–3139.

Bronkhorst, A. & Plomp, R., 1988. The effect of head-induced interaural time and level differences on speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 83(4), pp.1508–1516.

Brungart, D. et al., 2001. Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, 110(5), pp.2527–2538.

Buss, E. et al., 2008. Multicenter US bilateral MED-EL cochlear implantation study: speech perception over the first year of use. *Ear and hearing*, 29(1), pp.20–32.

Calvert, G., Bullmore, E. & Brammer, M., 1997. Activation of auditory cortex during silent lipreading. *Science*, 276(April), pp.593–595.

Carhart, R., Tillman, T. & Johnson, K., 1967. Release of masking for speech through interaural time delay. *The journal of the acoustical …*, 42(1), pp.124–138.

Cherry, E., 1953. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25(5), pp.975–979.

De Cheveigné, A., 1997. Concurrent vowel identification . III . A neural model. *The Journal of the Acoustical Society of America*, 101(5), pp.2857–2865.

De Cheveigné, A. & McAdams, S., 1995. Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement. *The Journal of the Acoustical Society of America*, 97(6), pp.3736–3748.

Ching, T., Dillon, H. & Byrne, D., 1998. Speech recognition of hearing-impaired listeners : Predictions from audibility and the limited role of high-frequency amplification. *The Journal of the Acoustical Society of America*, 103(2), pp.1128–1140.

Churchill, T. et al., 2014. Spatial hearing benefits demonstrated with presentation of acoustic temporal fine structure cues in bilateral cochlear implant listeners. *The Journal of the Acoustical Society of America*, 136(3), pp.1246–1256.

Clark, G., 2003. Cochlear implants in children: safety as well as speech and language. *International Journal of Pediatric Otorhinolaryngology*, 67(S1), pp.S7–S20.

Colburn, H., 1977. Erratum:"Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise". *The Journal of the Acoustical Society of America*, 62(5), p.1315.

Colburn, H., 1973. Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination. *The Journal of the Acoustical Society of America*, 54(6), pp.1458–1470.

Colburn, H. & Durlach, N., 1978. Models of binaural interaction. In E. Carterette & M. Friedman, eds. *Handbook of perception*. Academic Press, pp. 467–520.

Culling, J., 2007. Evidence specifically favoring the equalization-cancellation theory of binaural unmasking. *The Journal of the Acoustical Society of America*, 122(5), pp.2803–2813.

Culling, J., 1996. Signal-processing software for teaching and research in psychoacoustics under UNIX and X-Windows. *Behavior Research Methods, Instruments, & Computers*, 28(3), pp.376–382.

Culling, J., 2011. Subcomponent cues in binaural unmasking. *The Journal of the Acoustical Society of America*, 129(6), pp.3846–3855.

Culling, J. et al., 2012. The benefit of bilateral versus unilateral cochlear implantation to speech intelligibility in noise. *Ear and hearing*, 33(6), pp.673–682.

Culling, J. & Colburn, H., 2000. Binaural sluggishness in the perception of tone sequences and speech in noise. *The Journal of the Acoustical Society of America*, 107(1), pp.517–527.

Culling, J., Colburn, H. & Spurchise, M., 2001. Interaural correlation sensitivity. *The Journal of the Acoustical Society of America*, 110(2), pp.1020–1029.

Culling, J. & Darwin, C., 1993. Perceptual separation of simultaneous vowels: Within and across-formant grouping by F0. *The Journal of the Acoustical Society of America*, 93(6), pp.3454–3567.

Culling, J., Edmonds, B. & Hodder, K., 2006. Speech perception from monaural and binaural information. *The Journal of the Acoustical Society of America*, 119(1), pp.559–565.

Culling, J., Hawley, M. & Litovsky, R., 2005. Erratum: The role head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources [J. Acoust. Soc. Am. 116, 1057 (2004)]. *The Journal of the Acoustical Society of America*, 118(1), p.552.

Culling, J., Hawley, M. & Litovsky, R., 2004. The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources. *The Journal of the Acoustical Society of America*, 116(2), pp.1057–1065.

Culling, J., Lavandier, M. & Jelfs, S., 2013. Predicting Binaural Speech Intelligibility in Architectural Acoustics. In J. Blauert, ed. *The Technology of Binaural Listening*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 427–447.

Culling, J. & Summerfield, Q., 1998. Measurements of the binaural temporal window using a detection task. *The Journal of the Acoustical Society of America*, 103(6), pp.3540–3553.

Culling, J. & Summerfield, Q., 1995. Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay. *The Journal of the Acoustical Society of America*, 98(2), pp.785–797.

Deroche, M. et al., 2014. Speech recognition against harmonic and inharmonic complexes: spectral dips and periodicity. *The Journal of the Acoustical Society of America*, 135(5), pp.2873–2884.

Deroche, M. & Culling, J., 2011. Voice segregation by difference in fundamental frequency: Evidence for harmonic cancellation. *The Journal of the Acoustical Society of America*, 130(5), pp.2855–2865.

Desai, S., Stickney, G. & Zeng, F., 2008. Auditory-visual speech perception in normal-hearing and cochlear-implant listeners. *The Journal of the Acoustical Society of America*, 123(1), pp.428–40.

Van Deun, L., Van Wieringen, A. & Wouters, J., 2010. Spatial speech perception benefits in young children with normal hearing and cochlear implants. *Ear and hearing*, 31(5), pp.702–713.

Dirks, D. & Wilson, R., 1969. The Effect of Spatially Separated Sound Sources on Speech Intelligibility. *Journal of Speech Language and Hearing Research*, 12(1), pp.5–38.

Doherty, K. & Turner, C., 1996. Use of a correlational method to estimate a listener ' s weighting function for speech. *The Journal of the Acoustical Society of America*, 100(6), pp.3769–3773.

Domnitz, R. & Colburn, H., 1976. Analysis of binaural detection models for dependence on interaural target parameters. *The Journal of the Acoustical Society of America*, 59(3), pp.598–601.

Dorman, M. et al., 2007. An electric frequency-to-place map for a cochlear implant patient with hearing in the nonimplanted ear. *Journal of the Association for Research in Otolaryngology*, 8(2), pp.234–240.

Drullman, R., Festen, J. & Plomp, R., 1994. Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America*, 95(2), pp.1053–1064.

Dubbelboer, F. & Houtgast, T., 2008. The concept of signal-to-noise ratio in the modulation domain and speech intelligibility. *The Journal of the Acoustical Society of America*, 124(6), pp.3937–3946.

Dubno, J., Dirks, D. & Morgan, D., 1984. Effects of age and mild hearing loss on speech recognition in noise. *The Journal of the Acoustical Society of America*, 76(July 1984), pp.87–96.

Duda, R. & Martens, W., 1998. Range dependence of the response of a spherical head model. *The Journal of the Acoustical Society of America*, 104(5), pp.3048–3058.

Dugal, R., Braida, L. & Durlach, N., 1980. Implications of previous research for the selection of frequency-gain characteristics. *Acoustical factors affecting hearing aid performance*, pp.379–403.

Durlach, N., 1972. Binaural signal detection-Equalization and cancellation theory. In *Foundations of modern auditory theory*. New York, Academic Press, Inc., pp. 371–462.

Durlach, N., 1963. Equalization and Cancellation Theory of Binaural Masking-Level Differences. *The Journal of the Acoustical Society of America*, 35(8), pp.1206–1218.

Durlach, N. et al., 1986. Interaural correlation discrimination: II. Relation to binaural unmasking. *The Journal of the Acoustical Society of America*, 79(5), pp.1548–1557.

Dyer, W., 1962. The masking of speech by high-and low-pass noise. *Rome air development center Griffiss AFB, N Y*, Jul(AD0283122).

Eapen, R. et al., 2009. Hearing-in-Noise Benefits After Bilateral Simultaneous Cochlear Implantation Continue to Improve 4 Years After Implantation. *Otology & neurotology*, 30, pp.153–159.

Edmonds, B. & Culling, J., 2005. The role of head-related time and level cues in the unmasking of speech in noise and competing speech. *Acta acustica*, 91(3), pp.546–553.

Elhilali, M., Chi, T. & Shamma, S., 2003. A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. *Speech Communication*, 41(2-3), pp.331–348.

Elliott, L., 1975. Temporal and masking phenomena in persons with sensorineural hearing loss. *International Journal of Audiology*, 14, pp.336–353.

Erber, N., 1969. Interaction of Audition and Vision in the Recognition of Oral Speech Stimuli. *Journal of Speech Language and Hearing Research*, 12(2), pp.423–425.

Festen, J. & Plomp, R., 1990. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, 88(4), pp.1725–1736.

Fletcher, H., 1952. The Perception of Speech Sounds by Deafened Persons. *The Journal of the Acoustical Society of America*, 24(5), pp.490–497.

Fletcher, H. & Galt, R., 1950. The perception of speech and its relation to telephony. *The Journal of the Acoustical Society of America*, 22(2), pp.89–151.

French, N. & Steinberg, J., 1947. Factors governing the intelligibility of speech sounds. *The Journal of the Acoustical Society of America*, 19(1), pp.90–119.

Gabriel, K., 1983. Binaural interaction in hearing impaired listeners. *PhD dissertation*.

Gabriel, K. & Colburn, H., 1981. Interaural correlation discrimination: I. Bandwidth and level dependence. *The Journal of the Acoustical Society of America*, 69(5), p.1394.

Gardner, W. & Martin, K., 1995. HRTF measurements of a KEMAR. *The Journal of the Acoustical Society of America*, 97(6), pp.3907–3908.

Giraud, A. et al., 2001. Cross-modal plasticity underpins language recovery after cochlear implantation. *Neuron*, 30(3), pp.657–663.

Glasberg, B. & Moore, B., 1990. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1-2), pp.103–138.

Glyde, H. et al., 2011. Problems hearing in noise in older adults: a review of spatial processing disorder. *Trends in amplification*, 15(3), pp.116–126.

Goh, W. et al., 2001. Audio-Visual Perception of Sinewave Speech in an Adult Cochlear Implant User: A Case Study. *Ear and Hearing*, 22(5), pp.412–419.

Grant, K., 1987. Encoding voice pitch for profoundly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 82(2), pp.423–432.

Grant, K. & Braida, L., 1991. Evaluating the articulation index for auditory-visual input. *The Journal of the Acoustical Society of America*, 89(6), pp.2952–2960.

Grant, K. & Seitz, P., 2000. The use of visible speech cues for improving auditory detection. *The Journal of the Acoustical Society of America*, 108(3), pp.1197–1208.

Grant, K. & Walden, B., 1996. Evaluating the articulation index for auditory–visual consonant recognition. *The Journal of the Acoustical Society of America*, 100(April 1995), pp.2415–2424.

Grant, K., Walden, B. & Seitz, P., 1998. Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America*, 103(5), pp.2677–2690.

Grantham, D. & Wightman, F., 1979. Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation. *The Journal of the Acoustical Society of America*, 65(6), pp.1509–1517.

Grantham, D. & Wightman, F., 1978. Detectability of varying interaural temporal differencesa). *The Journal of the Acoustical Society of America*, 63(2), pp.511–523.

Hall, J. & Grose, J., 1988. Comodulation masking release: Evidence for multiple cues. *The Journal of the Acoustical Society of America*, 84(5), pp.1669–1675.

Hall, J., Tyler, R. & Fernandes, M., 1983. Monaural and binaural auditory frequency resolution measured using bandlimited noise and notched-noise masking. *The Journal of the Acoustical Society of America*, 73(3), pp.894–898.

Hawley, M., Litovsky, R. & Culling, J., 2004. The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *The Journal of the Acoustical Society of America*, 115(2), pp.833–843.

Hay-McCutcheon, M., Pisoni, D. & Kirk, K., 2005. Audiovisual speech perception in elderly cochlear implant recipients. *The Laryngoscope*, 115(10), pp.1887–1894.

Helfer, K., Chevalier, J. & Freyman, R., 2010. Aging, spatial cues, and single- versus dual-task performance in competing speech perception. *The Journal of the Acoustical Society of America*, 128(6), pp.3625–3633.

Helfer, K. & Freyman, R., 2005. The role of visual speech cues in reducing energetic and informational masking. *The Journal of the Acoustical Society of America*, 117(2), pp.842–849.

Henry, B. et al., 2000. The relationship between speech perception and electrode discrimination in cochlear implantees. *The Journal of the Acoustical Society of America*, 108(3 Pt.1), pp.1271–1280.

Hirsh, I., 1950. The relation between localization and intelligibility. *The Journal of the Acoustical Society of America*, 22(2), pp.196–200.

Van Hoesel, R., 2015. Audio-visual speech intelligibility benefits with bilateral cochlear implants when talker location varies dynamically. *Journal of the Association for Research in Otolaryngology : JARO*, 16, pp.309–315.

Van Hoesel, R. et al., 2008. Binaural speech unmasking and localization in noise with bilateral cochlear implants using envelope and fine-timing based strategies. *The Journal of the Acoustical Society of America*, 123(4), pp.2249–2263.

Van Hoesel, R., 2004. Exploring the benefits of bilateral cochlear implants. *Audiology and Neurotology*, 9, pp.234–246.

Van Hoesel, R., Jones, G. & Litovsky, R., 2009. Interaural time-delay sensitivity in bilateral cochlear implant users: effects of pulse rate, modulation rate, and place of stimulation. *Journal of the Association for Research in Otolaryngology*, 10(4), pp.557–567.

Van Hoesel, R. & Tyler, R., 2003. Speech perception, localization, and lateralization with bilateral cochlear implants. *The Journal of the Acoustical Society of America*, 113(3), pp.1617–1630.

Hohmann, V., 2002. Frequency analysis and synthesis using a Gammatone filterbank. *Acta Acustica United with Acustica*, 88(3), pp.433–442.

Holube, I., Kinkel, M. & Kollmeier, B., 1998. Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments. *The Journal of the Acoustical Society of America*, 104(4), pp.2412–2425.

Hopkins, K. & Moore, B., 2011. The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *The Journal of the Acoustical Society of America*, 130(1), pp.334–349.

Houtgast, T. & Steeneken, H., 1985. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *The Journal of the Acoustical Society of America*, 77(3), pp.1069–1077.

Vom Hövel, H., 1984. *Zur bedeutung der übertragungseigenschaften des außenohrs sowie des binauralen hörsystems bei gestörter sprach bertragng*. Fakultät für Elektrotechnik, RTWH Aachen.

Jain, M. et al., 1991. Fringed correlation discrimination and binaural detection. *The Journal of the Acoustical Society of America*, 90(4), pp.1918–1926.

Jeffress, L., 1948. A place theory of sound localization. *Journal of comparative and physiological psychology*, 41(1), pp.35–39.

Jelfs, S., 2011. Modelling the Cocktail Party : A Binaural Model for Speech Intelligibility in Noise. *PhD dissertation*.

Jelfs, S., Culling, J. & Lavandier, M., 2011. Revision and validation of a binaural model for speech intelligibility in noise. *Hearing research*, 275(1-2), pp.96–104.

Jørgensen, S. & Dau, T., 2013. Modelling Speech Intelligibility in Adverse Conditions. In B. Moore et al., eds. *Basic Aspects of Hearing*. 2012 ISH Proceedings, Springer New York, pp. 343–351.

Jørgensen, S. & Dau, T., 2011. Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. *The Journal of the Acoustical Society of America*, 130(3), pp.1475–1487.

Kalikow, D., Stevens, K. & Elliott, L., 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, pp.1337–1351.

Kendon, A., 1967. Some functions of gaze-direction in social interaction. *Acta psychologica*, 26, pp.22–63.

Kiang, N., 1968. A survey of recent developments in the study of auditory physiology. *The Annals of otology, rhinology, and laryngology*, 77(4), p.656.

Kiang, N., 1965. *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve.*,

King, A., Hopkins, K. & Plack, C., 2014. The effects of age and hearing loss on interaural phase difference discrimination. *The Journal of the Acoustical Society of America*, 135(1), pp.342–351.

Kock, W., 1950. Binaural localization and masking. *The Journal of the Acoustical Society of America*, 22(6), pp.801–804.

Koehnke, J. & Besing, J., 1996. A procedure for testing speech intelligibility in a virtual listening environment. *Ear and Hearing*, 17(3), pp.211–217.

Koenig, W., 1950. Subjective effects in binaural hearing. *The Journal of the Acoustical Society of America*, 22(1), pp.1–2.

Kollmeier, B. & Holube, I., 1992. Auditory filter bandwidths in binaural and monaural listening conditions. *The Journal of the Acoustical Society of America*, 92(4, Pt.1), pp.1889–1901.

Kryter, K., 1962a. Methods for the calculation and use of the articulation index. *The Journal of the Acoustical Society of America*, 34(11), pp.1689–1697.

Kryter, K., 1962b. Validation of the articulation index. *The Journal of the Acoustical Society of America*, 34(11), pp.1698–1702.

Kuhn, G., 1977. Model for the interaural time differences in the azimuthal plane. *The Journal of the Acoustical Society of America*, 62(1), pp.157–167.

Lachs, L., Pisoni, D. & Kirk, K., 2001. Use of Audiovisual Information in Speech Perception by Prelingually Deaf Children with Cochlear Implants: A First Report. *Ear and Hearing*, 22(3), pp.236–251.

Landry, S. et al., 2012. Audiovisual Segregation in Cochlear Implant Users. *PlosOne*, 7(3), pp.7–10.

Laske, R. et al., 2009. Subjective and objective results after bilateral cochlear implantation in adults. *Otology & Neurotology*, 30(3), pp.313–318.

Laszig, R. et al., 2004. Benefits of bilateral electrical stimulation with the nucleus cochlear implant in adults: 6-month postoperative results. *Otology & Neurotology*, 25(6), pp.958–68.

Lavandier, M. et al., 2012. Binaural prediction of speech intelligibility in reverberant rooms with multiple noise sources. *The Journal of the Acoustical Society of America*, 131(1), pp.218–231.

Lavandier, M. & Culling, J., 2010. Prediction of binaural speech intelligibility against noise in rooms. *The Journal of the Acoustical Society of America*, 127(1), pp.387–399.

Levitt, H. & Rabiner, L., 1967a. Binaural release from masking for speech and gain in intelligibility. *The Journal of the Acoustical Society of America*, 42(3), pp.601–608.

Levitt, H. & Rabiner, L., 1967b. Predicting binaural gain in intelligibility and release from masking for speech. *The Journal of the Acoustical Society of America*, 42(4), pp.820–829.

LeZak, R., Siegenthaler, B. & Davies, A., 1964. Bekesy-type audiometry for speech reception threshold. *Journal of Auditory Research*, 4(3), pp.181–189.

Licklider, J., 1948. The influence of interaural phase relations upon the masking of speech by white noise. *The Journal of the Acoustical Society of America*, 20(2), pp.150–159.

Lindemann, W., 1986a. Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. *The Journal of the Acoustical Society of America*, 80(6), pp.1608–1622.

Lindemann, W., 1986b. Extension of a Binaural Cross-Correlation Model by Means of Contralateral Inhibition. II. The Law of the First Wave Front. *Journ. Acoust. Soc. Am.*, 80(6), pp.1623–1630.

Litovsky, R. et al., 2010. Effect of age at onset of deafness on binaural sensitivity in electric hearing in humans. *The Journal of the Acoustical Society of America*, 127(1), pp.400–414.

Litovsky, R. et al., 2006. Simultaneous Bilateral Cochlear Implantation in Adults: A Multicenter Clinical Study. *Ear and Hearing*, 27(6), pp.714–731.

Litovsky, R., Parkinson, A. & Arcaroli, J., 2009. Spatial hearing and speech intelligibility in bilateral cochlear implant users. *Ear and hearing*, 30(4), pp.419–431.

Loizou, P. et al., 2009. Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *The Journal of the Acoustical Society of America*, 125(1), pp.372–383.

Long, C. et al., 2003. Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user. *The Journal of the Acoustical Society of America*, 114(3), pp.1565–1574.

Lovett, R. et al., 2010. Bilateral or unilateral cochlear implantation for deaf children: an observational study. *Archives of disease in childhood*, 95(2), pp.107–12.

Ludvigsen, C. et al., 1990. Prediction of intelligibility of non-linearly processed speech. *Acta oto-laryngologica. Supplementum*, (469), pp.190–195.

Lutfi, R., 1995. Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks. *The Journal of the Acoustical Society of America*, 97(2), pp.1333–1334.

Macleod, A. & Summerfield, Q., 1990. A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *British Journal of Audiology*, 24, pp.29–43.

MacLeod, A. & Summerfield, Q., 1987. Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, 21, pp.131–141.

Marrone, N., Mason, C. & Kidd, G., 2008a. The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. *The Journal of the Acoustical Society of America*, 124(5), pp.3064–3075.

Marrone, N., Mason, C. & Kidd, G., 2008b. Tuning in the spatial dimension: evidence from a masked speech identification task. *The Journal of the Acoustical Society of America*, 124(2), pp.1146–1158.

Martin, W., 1931. Rating the Transmission Performance of Telephone Circuits. *Bell System Technical Journal*, 10(1), pp.116–131.

Massaro, D., 1998. Perceiving talking faces: From speech perception to a behavioral principle. *American Scientist*.

McGrath, M. & Summerfield, Q., 1985. Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *The Journal of the Acoustical Society of America*, 77(2), pp.678–685.

McGurk, H. & MacDonald, J., 1976. Hearing lips and seeing voices. *Nature*, 264(5588), pp.746–748.

Meddis, R., 1988. Simulation of auditory-neural Re uptake. *The Journal of the Acoustical Society of America*, 83(3), pp.1056–1063.

Meddis, R., 1986. Simulation of mechanical to neural transduction in the auditory receptor. *The Journal of the Acoustical Society of America*, 79(3), pp.702–711.

Meddis, R., Hewitt, M. & Shackleton, T., 1990. Implementation details of a computation model of the inner hair-cell/auditory-nerve synapse. *The Journal of the Acoustical Society of America*, 87(4), pp.1813–1816.

Mehr, M., Turner, C. & Parkinson, A., 2001. Channel weights for speech recognition in cochlear implant users. *The Journal of the Acoustical Society of America*, 109(1), pp.359–366.

Meredith, M., Nemitz, J. & Stein, B., 1987. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of neuroscience*, 7(10), pp.3215–3229.

Middelweerd, M. & Plomp, R., 1987. The effect of speechreading on the speech-reception threshold of sentences in noise. *The Journal of the Acoustical Society of America*, 82(6), pp.2145–2147.

Miller, G., 1947. The masking of speech. *Psychological Bulletin*, 44(2), pp.105–129.

Miller, G. & Licklider, J., 1950. The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America*, 22(2), pp.167–173.

Miller, G. & Nicely, P., 1955. An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, 27(2), pp.338–352.

Misurelli, S. & Litovsky, R., 2012. Spatial release from masking in children with normal hearing and with bilateral cochlear implants: Effect of interferer asymmetry. *The Journal of the Acoustical Society of America*, 132(1), pp.380–391.

Mitchell, T. & Maslin, M., 2007. How vision matters for individuals with hearing loss. *International journal of audiology*, 46(9), pp.500–511.

Moody-Antonio, S. & Takayanagi, S., 2005. Improved speech perception in adult congenitally deafened cochlear implant recipients. *Otology & neurotology*, 26(4), pp.649–654.

Moore, B., 2012. *An Introduction to the Psychology of Hearing*, Moore, B., 2002. Response to “Articulation index predictions for hearing-impaired listeners with and without cochlear dead regions” [J. Acoust. Soc. Am. 111, 2545–2548 (2002)]. *The Journal of the Acoustical Society of America*, 111(6), p.2549.

Moore, B. et al., 2012. The influence of age and high-frequency hearing loss on sensitivity to temporal fine structure at low frequencies (L). *The Journal of the Acoustical Society of America*, 131(2), pp.1003–1006.

Moore, B. & Glasberg, B., 1983. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, 74(September), pp.750–753.

Moore, B. & Glasberg, B., 1998. Use of a Loudness Model for Hearing-Aid Fitting. I. Linear Hearing Aids. *British Journal of Audiology*, 32(5), pp.317–335.

Muir, D. & Field, J., 1979. Newborn Infants Orient to Sounds. *Child Development*, 50(2), pp.431–435.

Müller, C., 1992. *Perzeptive Analyse und Weiterentwicklung eines Reimtestverfahrens für die Sprachaudiometrie*.

Müller, J., Schon, F. & Helms, J., 2002. Speech Understanding in Quiet and Noise in Bilateral Users of the MED-EL COMBI 40/40+ Cochlear Implant System. *Ear and Hearing*, 23(3), pp.198–206.

Munson, W., 1945. Relation between the theory of hearing and the interpretation of speech sounds. *The Journal of the Acoustical Society of America*, 17, p.103.

Murphy, J. et al., 2011. Spatial hearing of normally hearing and cochlear implanted children. *International journal of pediatric otorhinolaryngology*, 75(4), pp.489–494.

Nittrouer, S. & Boothroyd, A., 1990. Context effects in phoneme and word recognition by young children and older adults. *The Journal of the Acoustical Society of America*, 87(6), pp.2705–2715.

Van de Par, S., Trahiotis, C. & Bernstein, L., 2001. A consideration of the normalization that is typically included in correlation-based models of binaural detection. *The Journal of the Acoustical Society of America*, 109(2), pp.830–833.

Peissig, J. & Kollmeier, B., 1997. Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. *The Journal of the Acoustical Society of America*, 101(3), pp.1660–1670.

Peters, B. et al., 2007. Importance of age and postimplantation experience on speech perception measures in children with sequential bilateral cochlear implants. *Otology & Neurotology*, 28(5), pp.649–657.

Platte, H. & vom Hövel, H., 1980. On the Interpretation of the Results of Speech Intelligibility Measurements in the Presence of Noise in a Free-Field. *Acustica*, 45(3), pp.139–150.

Plomp, R., 1986. A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech and Hearing Research*, 29, pp.146–154.

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. *Acustica*, 34(4), pp.200–211.

Plomp, R. & 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, pp.326–328.

Plomp, R. & Mimpen, A., 1979a. Improving the reliability of testing the speech reception threshold for sentences. *International Journal of Audiology*, 18(1), pp.43–52.

Plomp, R. & Mimpen, A., 1979b. Speech-reception threshold for sentences as a function of age and noise level. *The Journal of the Acoustical Society of America*, 66(5), pp.1333–1342.

Pollack, I., 1948. Effects of high pass and low pass filtering on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America*, 20(3), pp.259–266.

Reiss, L., Lowder, M. & Karsten, S., 2011. Effects of extreme tonotopic mismatches between bilateral cochlear implants on electric pitch perception: A case study. *Ear and hearing*, 32(4), pp.536–540.

Repp, B., Frost, R. & Zsiga, E., 1992. Lexical mediation between sight and sound in speechreading. *The Quarterly Journal of Experimental Psychology*, 45A(1), pp.1–20.

Rhebergen, K. & Versfeld, N., 2005. A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *The Journal of the Acoustical Society of America*, 117(4), pp.2181–2192.

Rhebergen, K., Versfeld, N. & Dreschler, W., 2006. Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise. *The Journal of the Acoustical Society of America*, 120(6), pp.3988–3997.

Rhebergen, K., Versfeld, N. & Dreschler, W., 2008. Prediction of the intelligibility for speech in real-life background noises for subjects with normal hearing. *Ear and hearing*, 29(2), pp.169–175.

Rosen, S., Fourcin, A. & Moore, B., 1981. Voice pitch as an aid to lipreading. *Nature*, 291(May), pp.150–152.

Rothauser, E., Chapman, W. & Guttman, N., 1969. IEEE recommended practice for speech quality measurements. *IEEE Transactions on audio and electro-acoustics*, 17(3), pp.225–246.

Rouger, J. et al., 2007. Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proceedings of the National Academy of Sciences of the United States of America*, 104(17), pp.7295–7300.

Sams, M. et al., 1991. Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, 127(1), pp.141–145.

Schleich, P., Nopp, P. & D'Haese, P., 2004. Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBI 40/40+ cochlear implant. *Ear and hearing*, 25(3), pp.197–204.

Schorr, E. et al., 2005. Auditory-visual fusion in speech perception in children with cochlear implants. *Proceedings of the National Academy of Sciences of the United States of America*, 102(51), pp.18748–18750.

Schroeder, M., 1965. New method of measuring reverberation time. *The Journal of the Acoustical Society of America*, 37, pp.409–412.

Schubert, E., 1956. Some preliminary experiments on binaural time delay and intelligibility. *The journal of the acoustical society of america*, 28(5), pp.895–901.

Siebert, W., 1970. Frequency discrimination in the auditory system: Place or periodicity mechanisms? *Proceedings of the IEEE*, 58(5), pp.723–730.

Siebert, W., 1968. Stimulus transformations in the peripheral auditory system. In K. P & E. M, eds. *Recognizing patterns*. MIT Press.

Steeneken, H. & Houtgast, T., 1980. A physical method for measuring speech-transmission quality. *The Journal of the Acoustical Society of America*, 67(1), pp.318–326.

Steeneken, H. & Houtgast, T., 1983. The temporal envelope spectrum and its significance in room acoustics. In *Proc. Int. Cong. Acoust.* pp. 85–88.

Strelnikov, K. et al., 2009. Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf adults with cochlear implants. *Scandinavian journal of psychology*, 50(5), pp.437–444.

Sumby, W. & Pollack, I., 1954. Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 26(2), pp.212–215.

Summerfield, Q., 1992. Lipreading and audio-visual speech perception. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, (335), pp.71–78.

Summerfield, Q., 1987. Some preliminaries to a comprehensive account of audio-visual speech perception. In *Hearing by Eye: the Psychology of Lip-Reading, edited by B. Dodd and R. Campbell (Lawrence Erlbaum, Hillsdale, NJ).*

Summerfield, Q. & McGrath, M., 1984. Detection and resolution of audio-visual incompatibility in the perception of vowels. *The Quarterly journal of experimental psychology. A, Human experimental psychology*, 36(1), pp.51–74.

Thurlow, W., Mangels, J. & Runge, P., 1967. Head movements during sound localization. *The Journal of the Acoustical Society of America*, 42(2), pp.489–493.

Tremblay, C. et al., 2010. Audiovisual fusion and cochlear implant proficiency. *Restorative neurology and neuroscience*, 28(2), pp.283–291.

Turner, C. et al., 1998. Frequency-weighting functions for broadband speech as estimated by a correlational method. *The Journal of the Acoustical Society of America*, 104(3), pp.1580–1585.

Tyler, R. et al., 2007. Speech perception and localization with adults with bilateral sequential cochlear implants. *Ear and Hearing*, 28(2 (supplement)), pp.865–905.

Tyler, R. et al., 2002. Three-Month Results with Bilateral Cochlear Implants. *Ear and Hearing*, 23(Supplement), p.80S–89S.

Vickers, D., Moore, B. & Baer, T., 2001. Effects of low-pass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies. *The Journal of the Acoustical Society of America*, 110(2), pp.1164–1175.

Walden, B.E., Prosek, R. a. & Worthington, D.W., 1975. Auditory and Audiovisual Feature Transmission in Hearing-Impaired Adults. *Journal of Speech Language and Hearing Research*, 18(2), pp.272–280.

Wallace, M., Meredith, M. & Stein, B., 1993. Converging Influences from Visual, Auditory, and Somatosensory Cortices Onto Output Neurons of the Smerior Colliculus. *J Neurophysiol*, 69(6), pp.1797–1809.

Wallach, H., 1940. The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, 27(4), pp.339–368.

Wan, R., Durlach, N. & Colburn, H., 2014. Application of a short-time version of the Equalization-Cancellation model to speech intelligibility experiments with speech maskers. *The Journal of the Acoustical Society of America*, 136(2), pp.768–776.

Wan, R., Durlach, N. & Colburn, H., 2010. Application of an extended equalization-cancellation model to speech intelligibility with spatially distributed maskers. *The Journal of the Acoustical Society of America*, 128(6), pp.3678–3690.

Webster, J. & Klumpp, R., 1963. Articulation Index and Average Curve-Fitting Methods of Predicting Speech Interference. *The Journal of the Acoustical Society of America*, 35(9), pp.1339–1344.

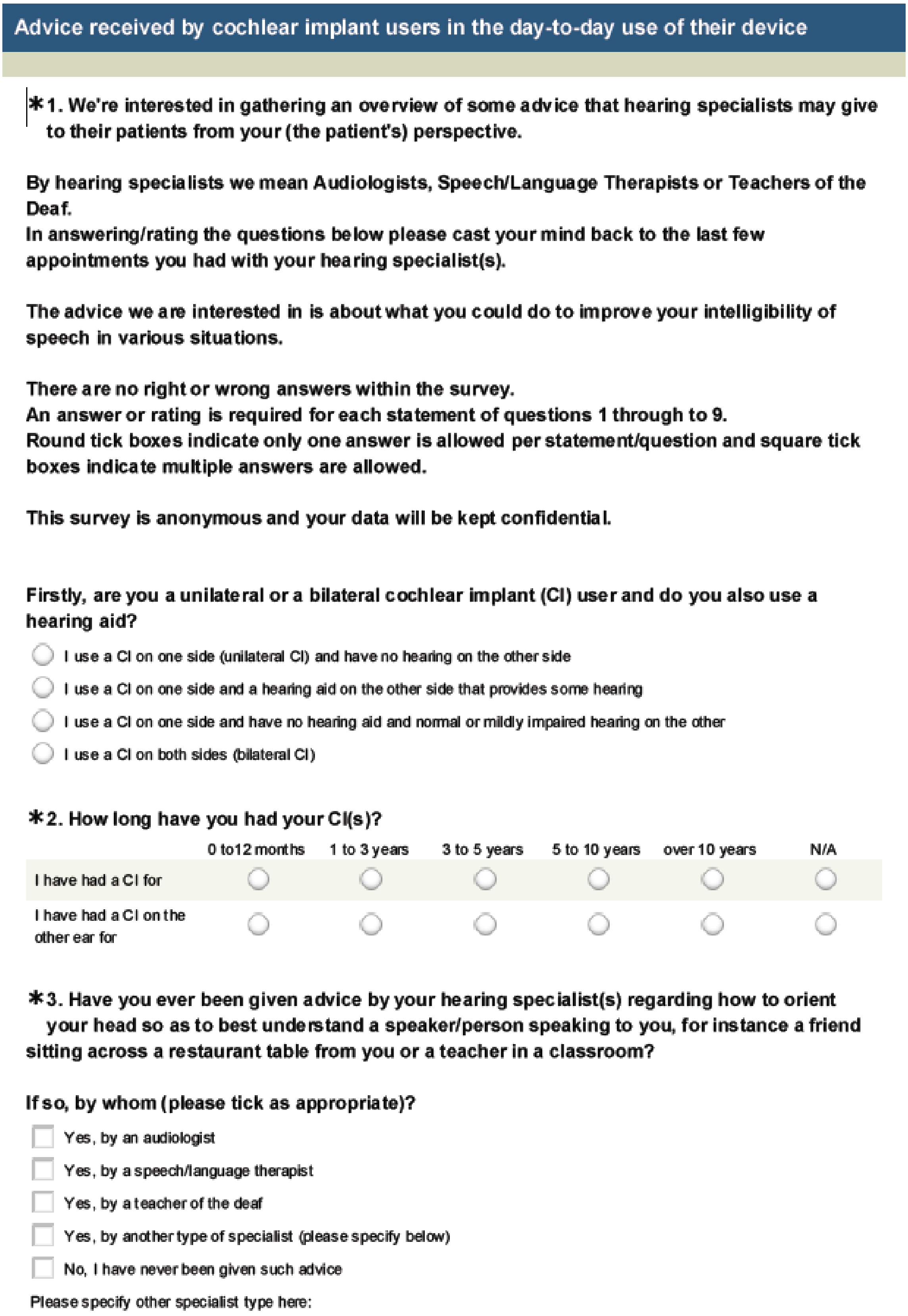
Weger, R. & Lane, C., 1924. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Physical Review*, 23(2), pp.266–285.

Wightman, F. & Kistler, D., 1999. Resolution of front-back ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America*, 105(5), pp.2841–53.

Young, P., 1931. The role of head movements in auditory localization. *Journal of Experimental Psychology*, 14(2), pp.95–124.

Zurek, P., 1993. Binaural advantages and directional effects in speech intelligibility. In G. Studebaker & I. Hochberg, eds. *Acoustical factors affecting hearing aid performance 2*. University Park Press, Baltimore, pp. 255–275.

APPENDIX 1: SURVEY OF CI USERS



Advice received by cochlear implant users in the day-to-day use of their device

*1. We're interested in gathering an overview of some advice that hearing specialists may give to their patients from your (the patient's) perspective.

By hearing specialists we mean Audiologists, Speech/Language Therapists or Teachers of the Deaf.
In answering/rating the questions below please cast your mind back to the last few appointments you had with your hearing specialist(s).

The advice we are interested in is about what you could do to improve your intelligibility of speech in various situations.

There are no right or wrong answers within the survey.
An answer or rating is required for each statement of questions 1 through to 9.
Round tick boxes indicate only one answer is allowed per statement/question and square tick boxes indicate multiple answers are allowed.

This survey is anonymous and your data will be kept confidential.

Firstly, are you a unilateral or a bilateral cochlear implant (CI) user and do you also use a hearing aid?

○ I use a CI on one side (unilateral CI) and have no hearing on the other side

○ I use a CI on one side and a hearing aid on the other side that provides some hearing

○ I use a CI on one side and have no hearing aid and normal or mildly impaired hearing on the other

○ I use a CI on both sides (bilateral CI)

*2. How long have you had your CI(s)?

	0 to12 months	1 to 3 years	3 to 5 years	5 to 10 years	over 10 years	N/A
I have had a CI for	○	○	○	○	○	○
I have had a CI on the other ear for	○	○	○	○	○	○

*3. Have you ever been given advice by your hearing specialist(s) regarding how to orient your head so as to best understand a speaker/person speaking to you, for instance a friend sitting across a restaurant table from you or a teacher in a classroom?

If so, by whom (please tick as appropriate)?

□ Yes, by an audiologist

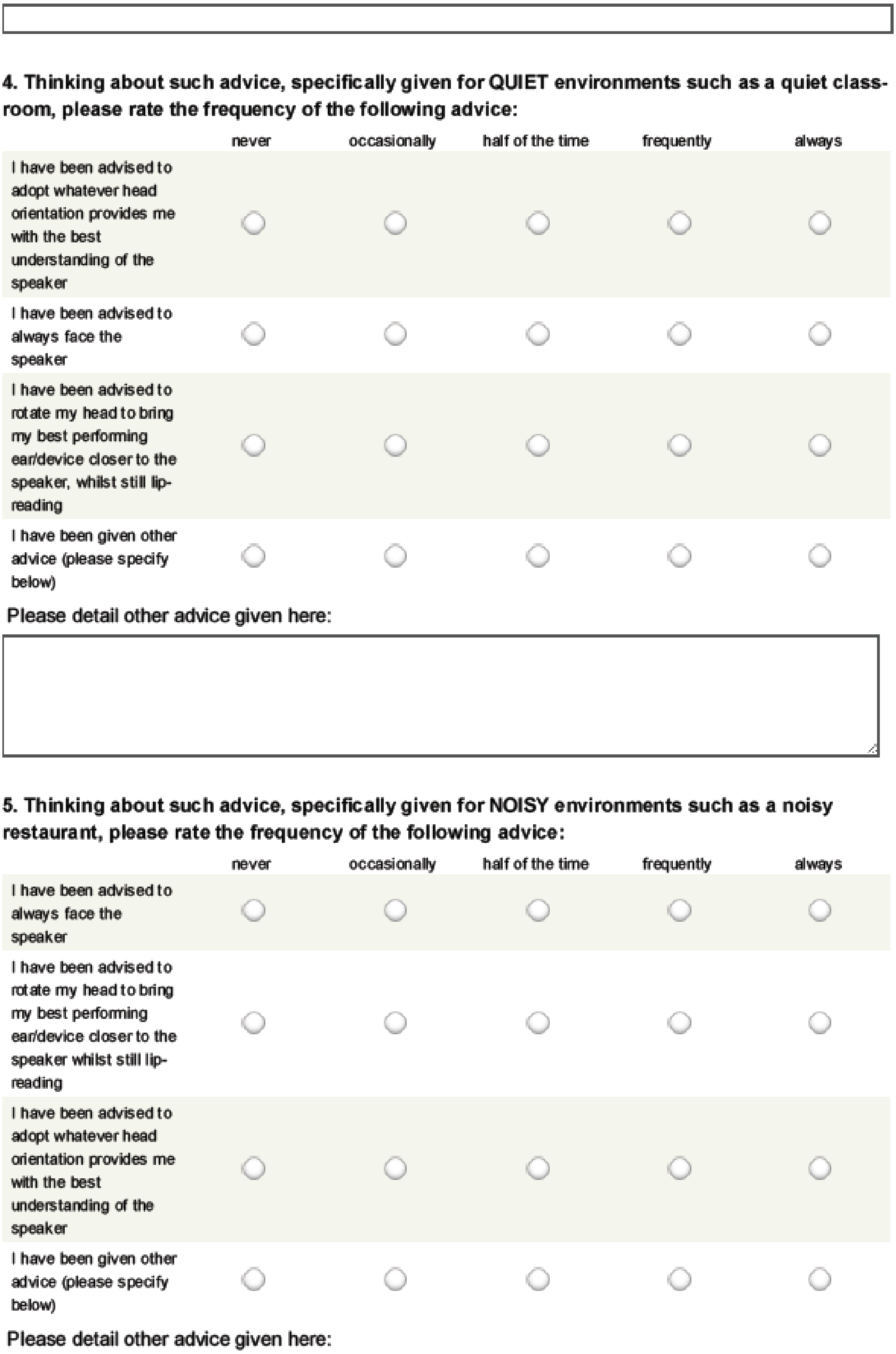
□ Yes, by a speech/language therapist

□ Yes, by a teacher of the deaf

□ Yes, by another type of specialist (please specify below)

□ No, I have never been given such advice

Please specify other specialist type here:



4. Thinking about such advice, specifically given for QUIET environments such as a quiet classroom, please rate the frequency of the following advice:

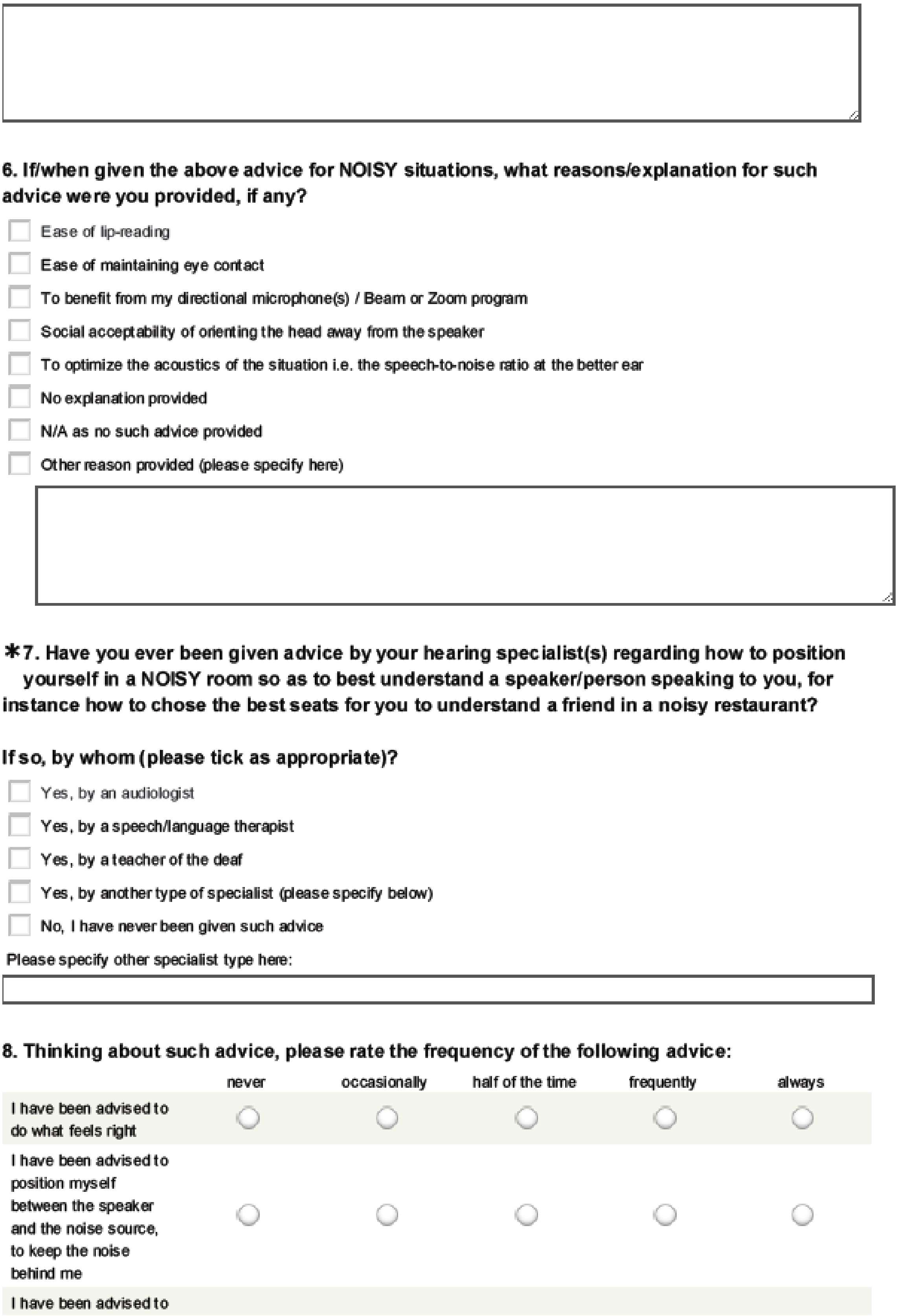
	never	occasionally	half of the time	frequently	always
I have been advised to adopt whatever head orientation provides me with the best understanding of the speaker	○	○	○	○	○
I have been advised to always face the speaker	○	○	○	○	○
I have been advised to rotate my head to bring my best performing ear/device closer to the speaker, whilst still lip-reading	○	○	○	○	○
I have been given other advice (please specify below)	○	○	○	○	○

Please detail other advice given here:

5. Thinking about such advice, specifically given for NOISY environments such as a noisy restaurant, please rate the frequency of the following advice:

	never	occasionally	half of the time	frequently	always
I have been advised to always face the speaker	○	○	○	○	○
I have been advised to rotate my head to bring my best performing ear/device closer to the speaker whilst still lip-reading	○	○	○	○	○
I have been advised to adopt whatever head orientation provides me with the best understanding of the speaker	○	○	○	○	○
I have been given other advice (please specify below)	○	○	○	○	○

Please detail other advice given here:



6. If/when given the above advice for NOISY situations, what reasons/explanation for such advice were you provided, if any?

□ Ease of lip-reading

□ Ease of maintaining eye contact

□ To benefit from my directional microphone(s) / Beam or Zoom program

□ Social acceptability of orienting the head away from the speaker

□ To optimize the acoustics of the situation i.e. the speech-to-noise ratio at the better ear

□ No explanation provided

□ N/A as no such advice provided

□ Other reason provided (please specify here)

***7. Have you ever been given advice by your hearing specialist(s) regarding how to position yourself in a NOISY room so as to best understand a speaker/person speaking to you, for instance how to chose the best seats for you to understand a friend in a noisy restaurant?**

If so, by whom (please tick as appropriate)?

□ Yes, by an audiologist

□ Yes, by a speech/language therapist

□ Yes, by a teacher of the deaf

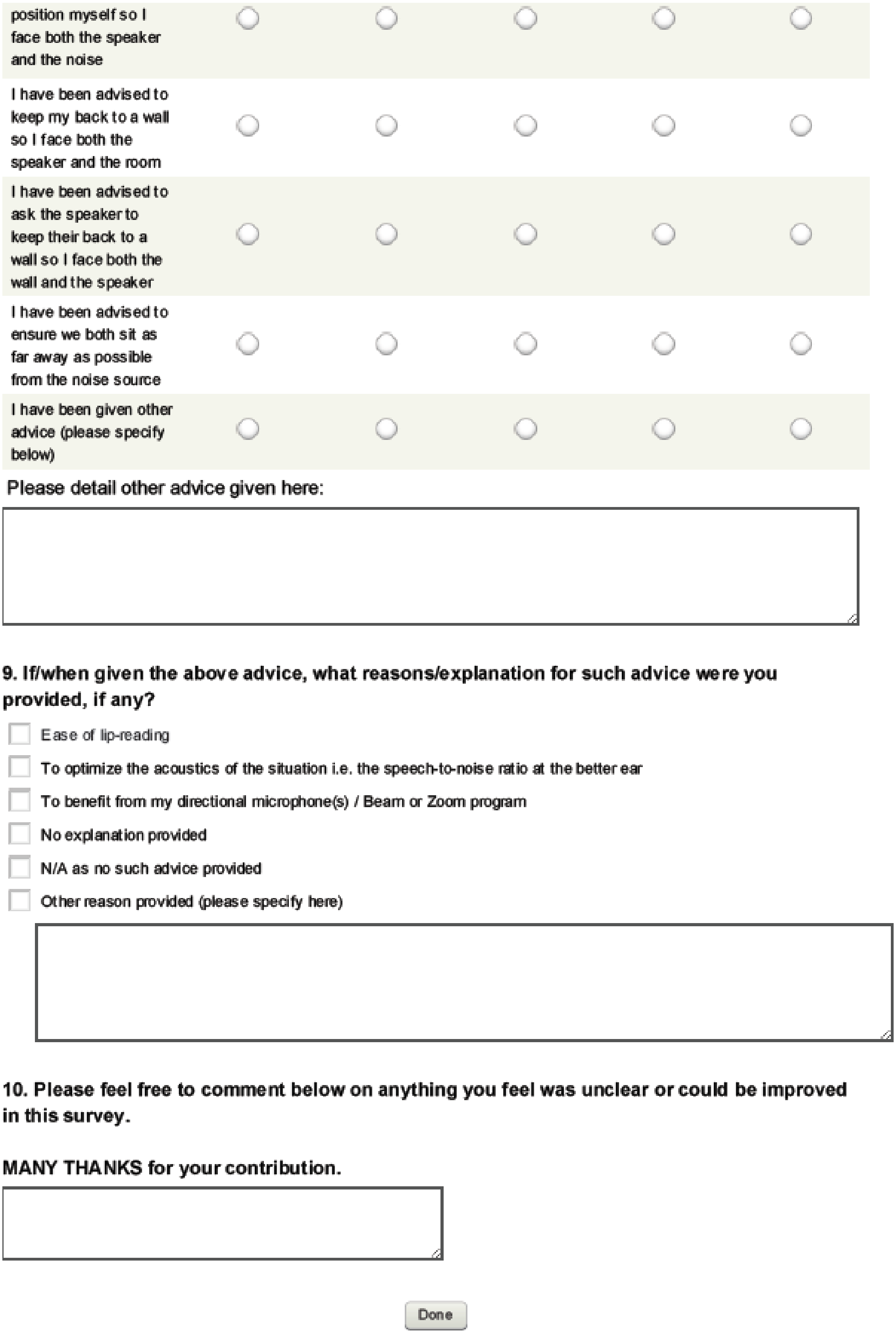
□ Yes, by another type of specialist (please specify below)

□ No, I have never been given such advice

Please specify other specialist type here:

8. Thinking about such advice, please rate the frequency of the following advice:

	never	occasionally	half of the time	frequently	always
I have been advised to do what feels right	○	○	○	○	○
I have been advised to position myself between the speaker and the noise source, to keep the noise behind me	○	○	○	○	○
I have been advised to					



position myself so I face both the speaker and the noise	○	○	○	○	○
I have been advised to keep my back to a wall so I face both the speaker and the room	○	○	○	○	○
I have been advised to ask the speaker to keep their back to a wall so I face both the wall and the speaker	○	○	○	○	○
I have been advised to ensure we both sit as far away as possible from the noise source	○	○	○	○	○
I have been given other advice (please specify below)	○	○	○	○	○

Please detail other advice given here:

9. If/when given the above advice, what reasons/explanation for such advice were you provided, if any?

□ Ease of lip-reading

□ To optimize the acoustics of the situation i.e. the speech-to-noise ratio at the better ear

□ To benefit from my directional microphone(s) / Beam or Zoom program

□ No explanation provided

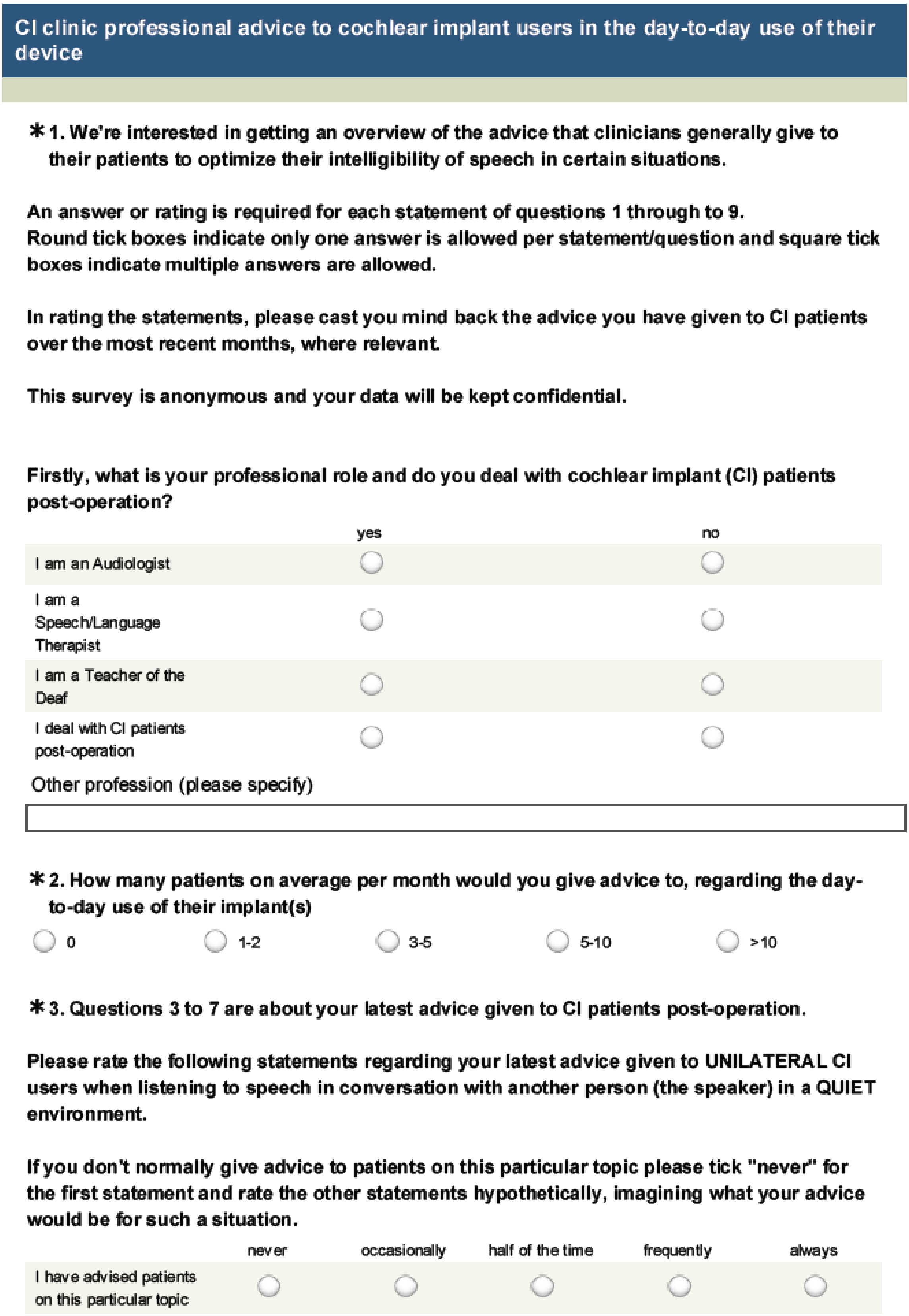
□ N/A as no such advice provided

□ Other reason provided (please specify here)

10. Please feel free to comment below on anything you feel was unclear or could be improved in this survey.

MANY THANKS for your contribution.

Appendix 2: Survey of CI Professionals



CI clinic professional advice to cochlear implant users in the day-to-day use of their device

*1. We're interested in getting an overview of the advice that clinicians generally give to their patients to optimize their intelligibility of speech in certain situations.

An answer or rating is required for each statement of questions 1 through to 9.
Round tick boxes indicate only one answer is allowed per statement/question and square tick boxes indicate multiple answers are allowed.

In rating the statements, please cast you mind back the advice you have given to CI patients over the most recent months, where relevant.

This survey is anonymous and your data will be kept confidential.

Firstly, what is your professional role and do you deal with cochlear implant (CI) patients post-operation?

	yes	no
I am an Audiologist	○	○
I am a Speech/Language Therapist	○	○
I am a Teacher of the Deaf	○	○
I deal with CI patients post-operation	○	○

Other profession (please specify)

□

*2. How many patients on average per month would you give advice to, regarding the day-to-day use of their implant(s)

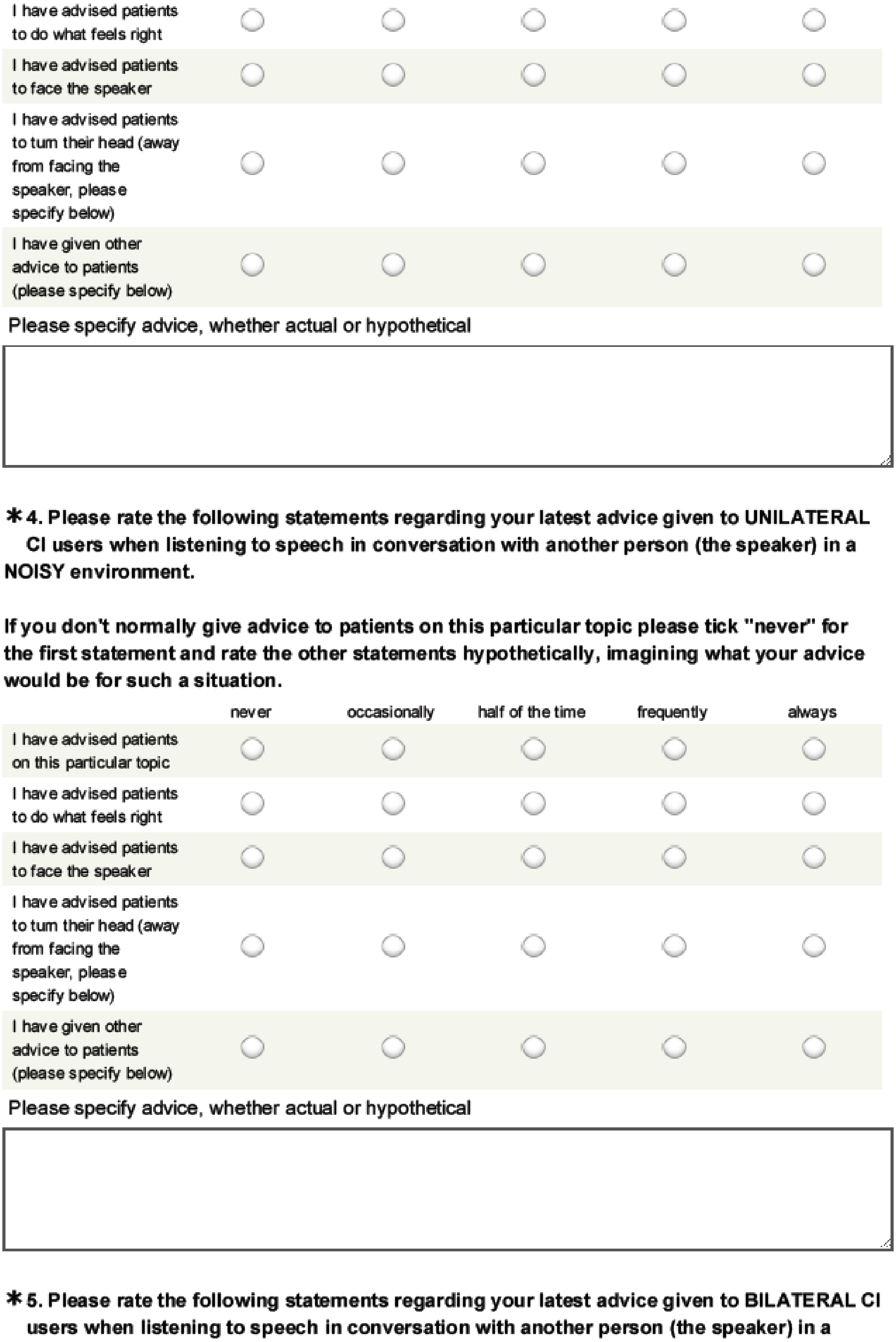
○ 0 ○ 1-2 ○ 3-5 ○ 5-10 ○ >10

*3. Questions 3 to 7 are about your latest advice given to CI patients post-operation.

Please rate the following statements regarding your latest advice given to UNILATERAL CI users when listening to speech in conversation with another person (the speaker) in a QUIET environment.

If you don't normally give advice to patients on this particular topic please tick "never" for the first statement and rate the other statements hypothetically, imagining what your advice would be for such a situation.

	never	occasionally	half of the time	frequently	always
I have advised patients on this particular topic	○	○	○	○	○



I have advised patients to do what feels right	○	○	○	○	○
I have advised patients to face the speaker	○	○	○	○	○
I have advised patients to turn their head (away from facing the speaker, please specify below)	○	○	○	○	○
I have given other advice to patients (please specify below)	○	○	○	○	○

Please specify advice, whether actual or hypothetical

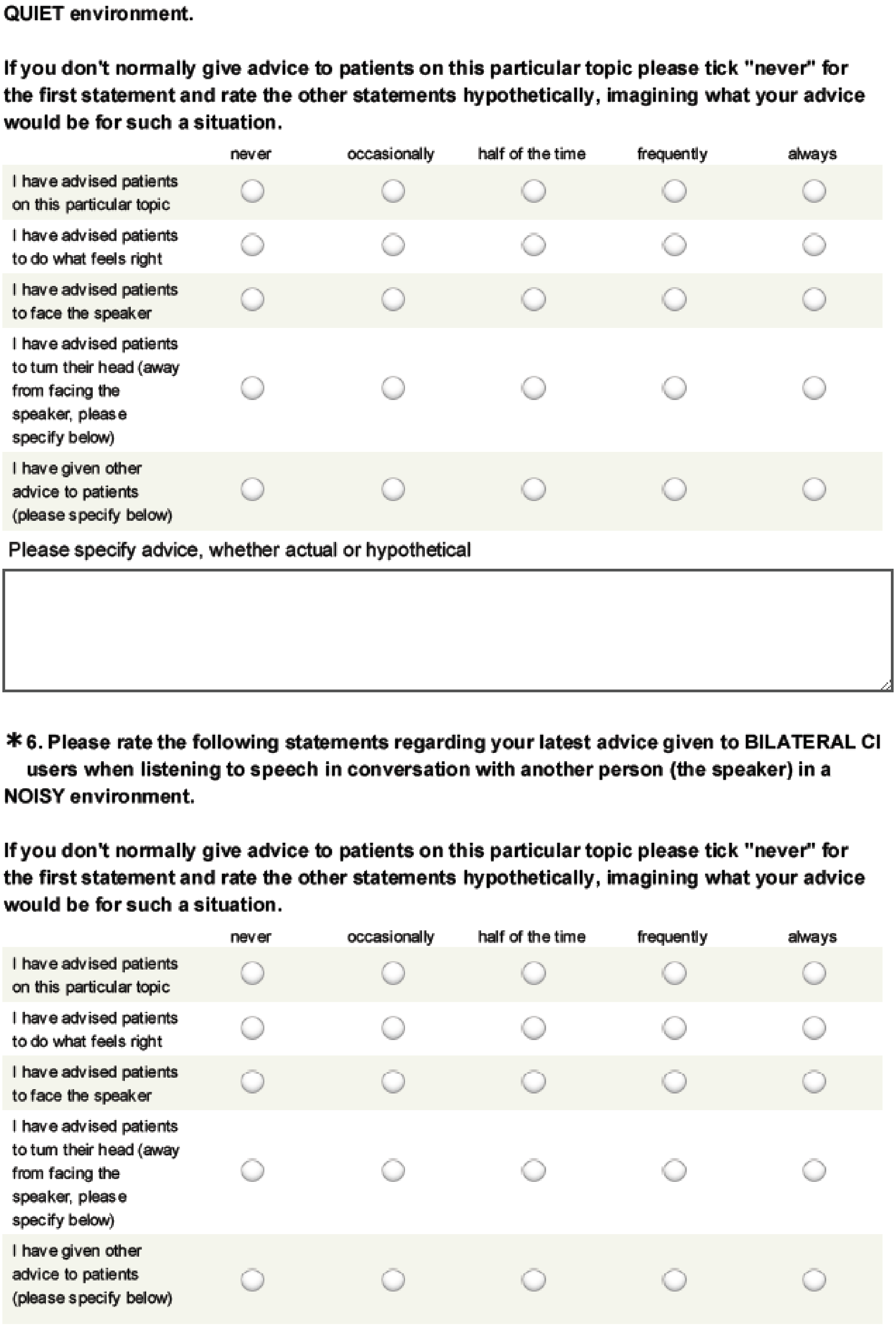
***4. Please rate the following statements regarding your latest advice given to UNILATERAL CI users when listening to speech in conversation with another person (the speaker) in a NOISY environment.**

If you don't normally give advice to patients on this particular topic please tick "never" for the first statement and rate the other statements hypothetically, imagining what your advice would be for such a situation.

	never	occasionally	half of the time	frequently	always
I have advised patients on this particular topic	○	○	○	○	○
I have advised patients to do what feels right	○	○	○	○	○
I have advised patients to face the speaker	○	○	○	○	○
I have advised patients to turn their head (away from facing the speaker, please specify below)	○	○	○	○	○
I have given other advice to patients (please specify below)	○	○	○	○	○

Please specify advice, whether actual or hypothetical

***5. Please rate the following statements regarding your latest advice given to BILATERAL CI users when listening to speech in conversation with another person (the speaker) in a**



QUIET environment.

If you don't normally give advice to patients on this particular topic please tick "never" for the first statement and rate the other statements hypothetically, imagining what your advice would be for such a situation.

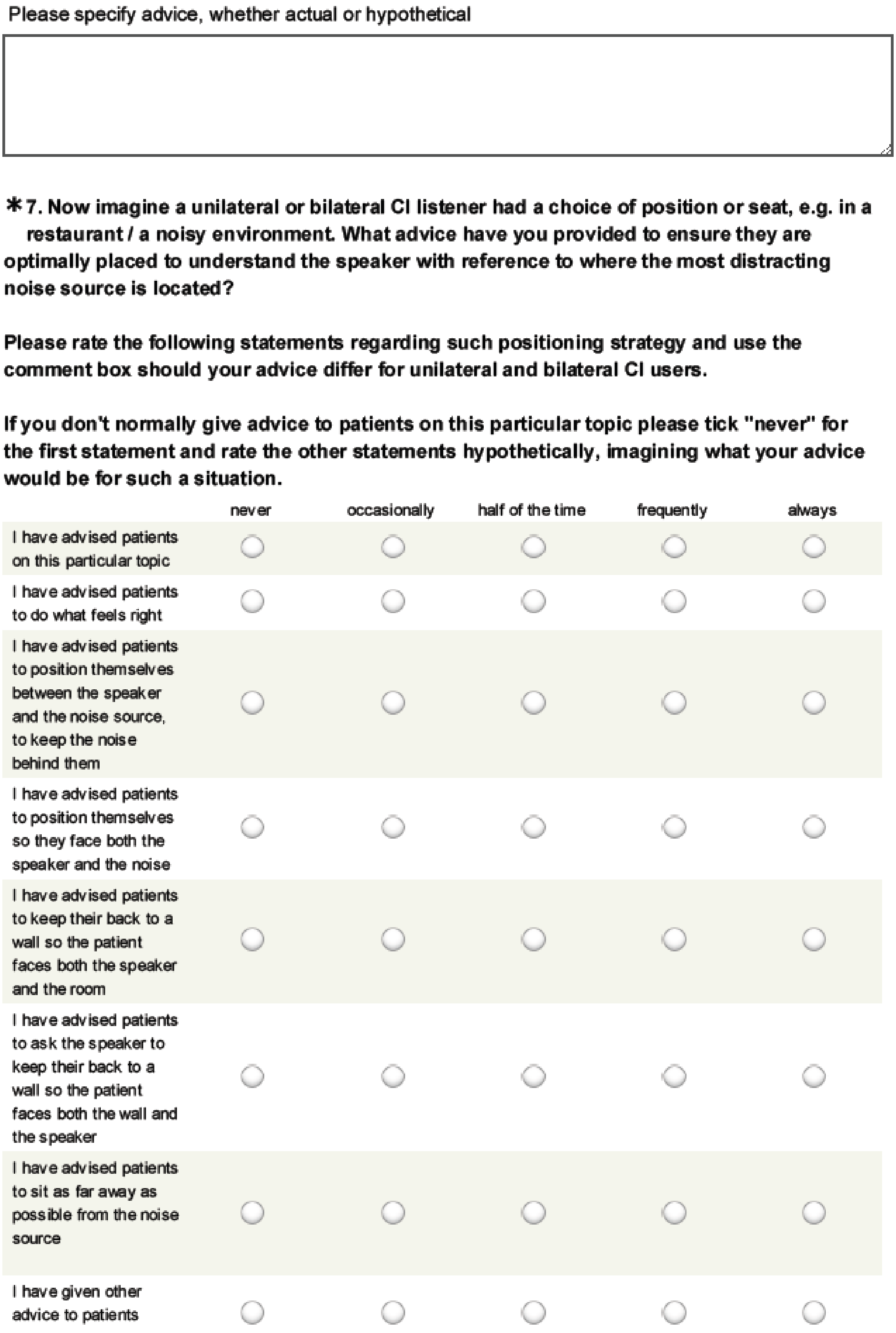
	never	occasionally	half of the time	frequently	always
I have advised patients on this particular topic	○	○	○	○	○
I have advised patients to do what feels right	○	○	○	○	○
I have advised patients to face the speaker	○	○	○	○	○
I have advised patients to turn their head (away from facing the speaker, please specify below)	○	○	○	○	○
I have given other advice to patients (please specify below)	○	○	○	○	○

Please specify advice, whether actual or hypothetical

*** 6. Please rate the following statements regarding your latest advice given to BILATERAL CI users when listening to speech in conversation with another person (the speaker) in a NOISY environment.**

If you don't normally give advice to patients on this particular topic please tick "never" for the first statement and rate the other statements hypothetically, imagining what your advice would be for such a situation.

	never	occasionally	half of the time	frequently	always
I have advised patients on this particular topic	○	○	○	○	○
I have advised patients to do what feels right	○	○	○	○	○
I have advised patients to face the speaker	○	○	○	○	○
I have advised patients to turn their head (away from facing the speaker, please specify below)	○	○	○	○	○
I have given other advice to patients (please specify below)	○	○	○	○	○



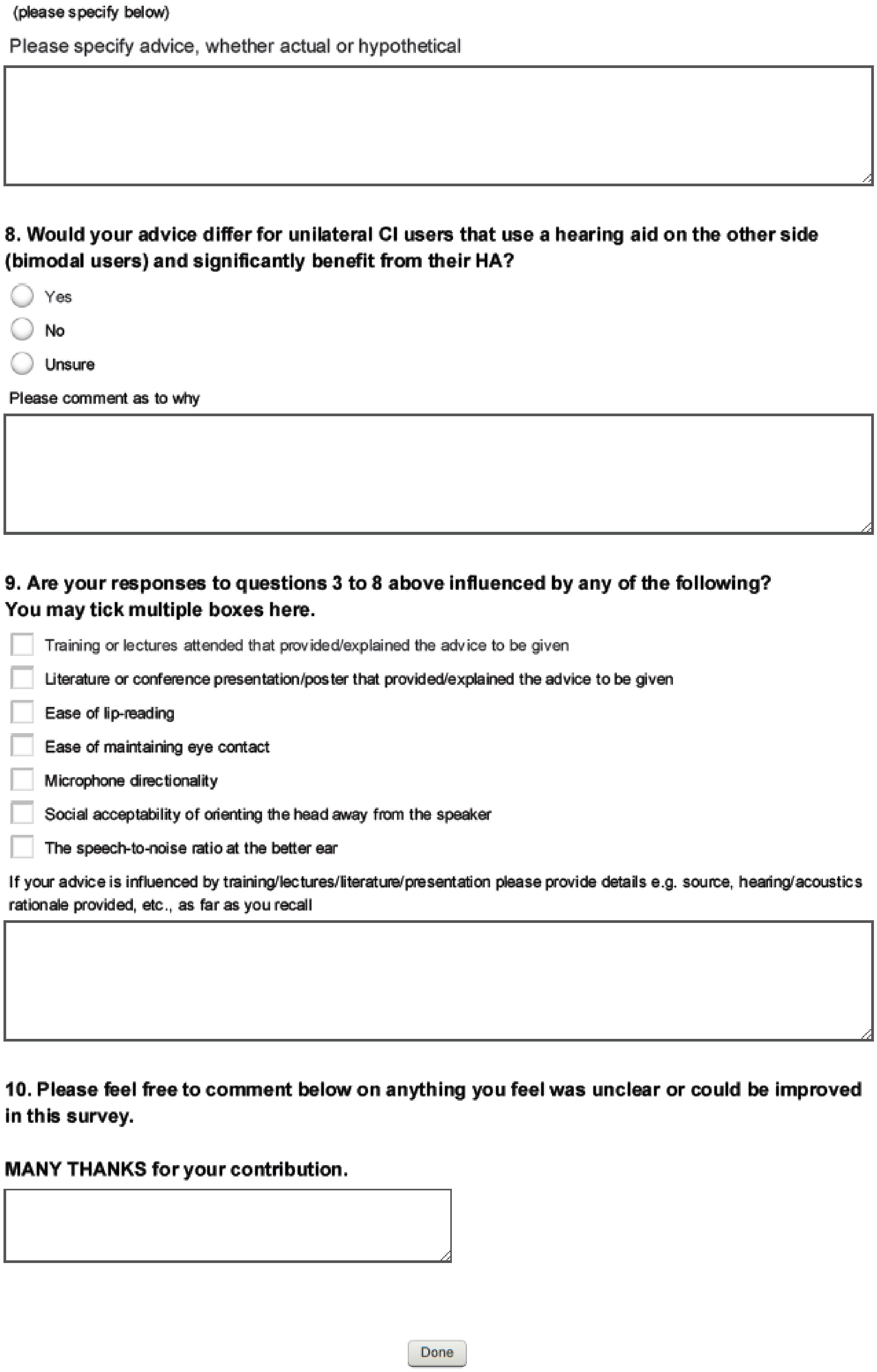
Please specify advice, whether actual or hypothetical

* 7. Now imagine a unilateral or bilateral CI listener had a choice of position or seat, e.g. in a restaurant / a noisy environment. What advice have you provided to ensure they are optimally placed to understand the speaker with reference to where the most distracting noise source is located?

Please rate the following statements regarding such positioning strategy and use the comment box should your advice differ for unilateral and bilateral CI users.

If you don't normally give advice to patients on this particular topic please tick "never" for the first statement and rate the other statements hypothetically, imagining what your advice would be for such a situation.

	never	occasionally	half of the time	frequently	always
I have advised patients on this particular topic	○	○	○	○	○
I have advised patients to do what feels right	○	○	○	○	○
I have advised patients to position themselves between the speaker and the noise source, to keep the noise behind them	○	○	○	○	○
I have advised patients to position themselves so they face both the speaker and the noise	○	○	○	○	○
I have advised patients to keep their back to a wall so the patient faces both the speaker and the room	○	○	○	○	○
I have advised patients to ask the speaker to keep their back to a wall so the patient faces both the wall and the speaker	○	○	○	○	○
I have advised patients to sit as far away as possible from the noise source	○	○	○	○	○
I have given other advice to patients	○	○	○	○	○



(please specify below)

Please specify advice, whether actual or hypothetical

8. Would your advice differ for unilateral CI users that use a hearing aid on the other side (bimodal users) and significantly benefit from their HA?

○ Yes

○ No

○ Unsure

Please comment as to why

9. Are your responses to questions 3 to 8 above influenced by any of the following? You may tick multiple boxes here.

□ Training or lectures attended that provided/explained the advice to be given

□ Literature or conference presentation/poster that provided/explained the advice to be given

□ Ease of lip-reading

□ Ease of maintaining eye contact

□ Microphone directionality

□ Social acceptability of orienting the head away from the speaker

□ The speech-to-noise ratio at the better ear

If your advice is influenced by training/lectures/literature/presentation please provide details e.g. source, hearing/acoustics rationale provided, etc., as far as you recall

10. Please feel free to comment below on anything you feel was unclear or could be improved in this survey.

MANY THANKS for your contribution.