

MACQUARIE UNIVERSITY

DOCTORAL THESIS

**Effect of noise and reverberation on speech
intelligibility with cochlear implants
considering realistic sound environments**

Author:

Javier BADAJOZ-DAVILA

Supervisors:

Dr. Jörg M. BUCHHOLZ

and

Dr. Richard VAN-HOESEL

*A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy*

in the

Department of Linguistics, Faculty of Human Sciences

February 11, 2020

Macquarie Univeristy
The Hearing Hub, 16 University Avenue, Macquarie University,
New South Wales 2109, Australia

© **Javier Badajoz-Davila 2019**

Effect of noise and reverberation on speech
intelligibility with cochlear implants
considering realistic sound environments
Javier Badajoz

Contents

Abstract	vii
Declaration of Authorship	ix
Acknowledgements	xiii
Chapter 1: Introduction	1
1.1 Background and motivation	1
1.1.1 Unrealistic reverberation	3
1.1.2 Unrealistic sound environments	4
1.1.3 Unrealistic speech material	4
1.1.4 Unrealistic SNRs	5
1.1.5 Unrealistic listening tasks	5
1.2 Approach	6
1.2.1 Realistic reverberation and sound environments	7
1.2.2 Realistic speech material	8
1.2.3 Realistic SNRs	9
1.3 Aims	9
1.4 Outline	10
Chapter 2: Effect of noise and reverberation on speech intelligibility for cochlear implant recipients in realistic sound environments	11
2.1 Introduction	12
2.2 Methods	14
2.2.1 Participants	14
2.2.2 Stimuli	14
2.2.2.1 Speech material	14
2.2.2.2 Sound reproduction	15
2.2.3 Cochlear Implant Research Platform (CIRP)	18
2.2.4 Procedures	20
2.2.5 Statistical analysis	21
2.3 Results	22
2.3.1 Speech intelligibility data	22
2.3.2 Questionnaire ratings	24
2.4 Discussion	25
2.4.1 Results in quiet	25

2.4.2	Results in noise	30
2.5	Conclusions	32
Chapter 3: Validation of existing room acoustic criteria for predicting speech intelligibility with cochlear implants		
		35
3.1	Introduction	36
3.2	Methods	39
3.2.1	Stimuli and speech intelligibility data	40
3.2.2	BTE signal simulation	42
3.2.3	Calculation of the U50	43
3.2.4	Temporal modulations of the noise	43
3.2.5	Statistical model	44
3.3	Results	46
3.3.1	Effect of reverberation in quiet	46
3.3.2	Effect of noise	48
3.3.3	Relative effects of noise and reverberation	50
3.4	Discussion	51
3.4.1	Speech intelligibility in quiet	51
3.4.1.1	Room parameters relevant to SI in quiet	53
3.4.1.2	Additional room acoustic factors	55
3.4.2	Speech intelligibility in noise	56
3.4.3	Speech Transmission Index (STI)	57
3.5	Outlook and limitations	58
3.6	Conclusions	59
Appendices		
3.A	Nonlinear mixed-effects model	61
3.B	Individual RMS errors	61
3.C	Relationship between the C50 and the STI	62
Chapter 4: Effect of test realism on speech-in-noise outcomes in bilateral cochlear implant users		
		65
4.1	Introduction	65
4.2	Methods	68
4.2.1	Subjects and speech processor	69
4.2.2	Stimuli	70
4.2.2.1	Acoustic scenes	70
4.2.2.2	Speech material	72
4.2.2.3	Sound reproduction	72
4.2.3	Procedures	73
4.2.4	Instrumental signal evaluation using the U50	74
4.2.5	Statistical analysis	75
4.3	Results	76

4.3.1	Effect of test realism on SI outcomes	77
4.3.2	Speech intelligibility in realistic conditions	82
4.3.3	Further analysis of bilateral SI advantage	83
4.4	Discussion	86
4.4.1	Effect of test realism on SI	86
4.4.2	Speech intelligibility in realistic conditions	89
4.5	Limitations and outlook	92
4.6	Conclusions	93
Appendices		
4.A	Biographic data of participants	95
4.B	Speech Reception Thresholds	96
Chapter 5:	Final considerations	97
5.1	Discussion	97
5.2	Outlook and limitations	101
5.3	Conclusions	104
Appendix A:	Higher Order Ambisonics	107
A.1	Sound field encoding	107
A.2	Shape matching	109
A.3	Sound field decoding	110
A.4	Regularization	111
Appendix B:	Ethics	113
Bibliography		119

Abstract

Understanding speech in background noise and reverberation is a known problem for individuals with cochlear implants (CIs). However, current laboratory-based assessments of speech intelligibility (SI) are usually poor predictors of an individual's listening abilities in the real-world. This mismatch is largely attributed to the use of oversimplified methods whereby neither the speech nor the noise employed in the tests resemble what an individual experiences in their daily life. In order to better understand the challenges faced by individuals listening with CIs, the present study employs materials and methods designed to bridge the gap between laboratory-based outcomes and individuals' experience of everyday listening.

This work comprises three main studies. The first two studies aim to systematically evaluate and understand the effect of realistic reverberation on SI in CI recipients. Sentence recall performance was measured in 12 unilateral CI recipients in both quiet and noise, considering six realistic rooms at varying target-to-receiver distances. The results suggest that in quiet conditions reverberation has a significant impact on SI mainly at long distances, with the exception of small reflective rooms, where SI is affected even at close, conversational distances. Further analysis of the data in quiet conditions suggests that room acoustic parameters such as the U50 can predict SI in rooms with reasonable accuracy. Analysis in noise revealed that the temporal smearing effect of reverberation on the noise signal is beneficial to SI, an effect that is not accounted for by the U50. Hence, future implementations of the U50 need to consider the noise-inherent modulations.

The goal of the third study is twofold. First, to understand the effect of test realism on SI outcomes. Second, to assess SI performance as well as bilateral benefit in CI recipients in more realistic noisy conditions. Sentence recall performance was measured in 15 bilateral CI recipients using sentence materials as well as noises with different level of realism. "Standard" BKB-like sentences were used as well as more realistic sentences that were cut out of natural two-talker conversations elicited at different vocal effort levels. Both sentence materials were presented in different realistic acoustic environments at natural signal-to-noise ratios (SNRs) as well as in "standard" babble noise. The results indicate that participants could more easily deal with babble noise than with more realistic noisy situations, and that they could understand more easily the standard sentences than more realistic (conversational) speech. This effect was pronounced at lower SNRs. A small but significant bilateral benefit was observed in most conditions.

The present work highlights the importance of using realistic reverberation, presentation levels, speech material, noise material and spatial representation of the sound field when assessing SI performance in CI recipients.

Declaration of Authorship

I, Javier BADAJOZ-DAVILA, declare that this thesis titled, “Effect of noise and reverberation on speech intelligibility with cochlear implants considering realistic sound environments” and the work presented in it are my own. I confirm that this work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Chapter 2 presents a study that was designed by Jörg Buchholz and myself. The Sydney Cochlear Implant Centre (SCIC) assisted in the recruitment of the participants. I carried out the measurements of the room impulse responses, with some assistance from Jörg Buchholz, and did the processing of the signals required to set up the experiment. I calibrated the xPC system and integrated it into the tests conducted in the anechoic chamber. I inserted and validated the fitting parameters of the subjects into the Simulink model. I was the experimenter during data collection. I conducted the data analyses with inputs and comments from Jörg Buchholz, Mark Seeto and Jaime Undurraga. I completed all the work, including plots and writing, with strong contributions from Jörg Buchholz and Richard Van Hoesel.

Chapter 3 further analyses the data presented in Chapter 2. I conducted calibrated simulations of the stimuli presented during the tests. I measured the required HATS impulse responses with assistance from James Galloway. I conducted all the statistical analyses included in this chapter with a rather significant contribution of Peter Humburg, who provided invaluable help in the approach adopted. Jörg Buchholz took an active part in the steps taken during the analyses, giving feedback and advice on the approaches taken. I wrote the chapter myself with feedback from Jörg Buchholz. The comparative analysis between the C50 and the STI was my idea and it was conducted by me.

Chapter 4 was jointly designed by Jörg Buchholz, Richard Van Hoesel and myself. The SCIC assisted in the recruitment of the participants. I conducted all the required work to set up the experiment including the acoustic scene selection, speech material auralization and calibration. Jörg Buchholz designed the Graphical User Interface to test the new speech material, and I integrated it into the multi-channel loudspeaker system. I programmed the speech processors with the participants’ fitting parameters. I conducted all the data collection. I conducted the BTE directivity microphone simulations and all the statistical analyses. Peter Humburg provided feedback on the statistical models and the correlation analysis. I wrote the chapter with strong contributions from Jörg Buchholz and Richard Van Hoesel.

Signed: _____ Date: 27/05/19
Sydney, Australia.

List of Publications

Conference posters and presentations

1. Badajoz-Davila, Javier, Jörg M. Buchholz & Richard Van Hoesel (2017). "Effect of reverberation on speech intelligibility in cochlear implant recipients considering realistic sound environments", *2017 Conference on Implantable Auditory Prostheses*, Lake Tahoe, US, July, 16th-21st.
2. Badajoz-Davila, Javier, Jörg M. Buchholz & Richard Van Hoesel (2018). "Effect of noise and reverberation on speech intelligibility in cochlear implant recipients considering realistic sound environments". *Audiology Australia National Conference 2018*, Sydney, Australia, May, 20th-23rd.
3. Badajoz-Davila, Javier, Jörg M. Buchholz & Richard Van Hoesel (2018). "Effect of noise and reverberation on speech intelligibility in cochlear implant recipients in realistic sound environments ". *15th international conference on cochlear implants and other implantable auditory technologies*, Antwerp, Belgium, June 27th-30th.
4. Badajoz-Davila, Javier, Jörg M. Buchholz & Richard Van Hoesel (2018). "Effect of noise and reverberation on speech intelligibility in cochlear implant recipients in realistic sound environments ". *Improving Cochlear Implant Performance 2018*, London, UK, July 24th.

To be submitted

5. Badajoz-Davila, Javier, Jörg M. Buchholz & Richard Van Hoesel. "Effect of noise and reverberation on speech intelligibility in cochlear implant recipients in realistic sound environments".
6. Badajoz-Davila, Javier, Jörg M. Buchholz & Richard Van Hoesel. "Validation of existing room acoustic criteria for predicting speech intelligibility with cochlear implants".
7. Badajoz-Davila, Javier, Jörg M. Buchholz & Richard Van Hoesel. "Effect of test realism on speech-in-noise outcomes in bilateral cochlear implant users".

Acknowledgements

I would like to thank my main supervisor Jörg Buchholz for his trust, his continuous support and for invaluable discussions. Thanks also to Richard Van Hoesel for interesting discussions and for his contributions. I would like to thank Robert Cowan for showing empathy and willingness to help during unforeseen circumstances. I wish to thank Michael Goorevich for his always-positive attitude towards this project, his continuous interest, support and for his effectiveness to assist whenever needed. Thanks also to Rachelle Hassarati and Dakota Bysouth-Young from the SCIC for being so supportive and effective with the recruitment of the participants.

The author acknowledges the financial support of the HEARing CRC, established under the Australian Government's Cooperative Research Centres (CRC) Program. The CRC Program supports industry-led collaborations between industry, researchers and the community. I would also like to acknowledge Macquarie University, Australia, for the scholarship provided.

During these three years, I was lucky enough to be surrounded by really helpful and kind people who helped me in a wide range of situations. Thanks to Kelly Miles and Timothy Beechey for all sorts of insightful conversations, for their help in statistics, data interpretation, writing as well as for keeping the motivation up with new and interesting points of view. Thanks to Peter Humburg for being so helpful, effective and clear. Thanks to Jaime Undurraga for his invaluable insights about hearing and statistics, for his eagerness to help and for his ability to actually help. Thanks to Greg Stewart for his help and for being able to always make it fun. Thanks as well to Katie Neal for helping with all the audiological considerations of the listening tests.

Of course, thanks to all the nice and fun people that made the everyday PhD lifestyle so much fun and relaxed. Thanks to Remi, Baljeet, Kiri, Kelly Miles and Timothy Beechey for our nonsense conversations. Thanks also to Joaquin, Fabrice and Bram for being so welcoming. Thanks to Adam Weisser for interesting and sometimes mind-blowing conversations. Thanks to my last office mates, Rakshita, Shivali and Isabelle for making this a nice working environment. I am also grateful for sharing office with *Hear for you*, whose invaluable support to teenagers with hearing difficulties makes me *want to be a better person*. I wish to thank all the members of the AHRG group for sharing their knowledge, for keeping up the good quality of research and for very fruitful discussions.

I also thank Denise for her generosity, which solved an uphill situation in a blink of an eye.

As always, thanks to my parents for their continuous support. And of course, thanks to Nans Liv Nielsen for being always there listening to the good and bad things and for making me feel that I was not doing this alone. Without her patience and ability to listen, I would probably not be here and I would probably be a gardener.

List of Abbreviations

APHAB	A bbreviated P rofile of H earing A id B enefit
ADRO	A utomatic D ynamic R ange O ptimisation
ANOVA	A Nalysis O f V ariance
ARTE	A mbisonic R ecordings of T ypical E nvironments
BKB	B amford, K owal and B ench
BTE	B ehind T he E ar
CI	C ochlear I mpant
CIRP	C ochlear I mpant R esearch P latform
DRR	D irect to R everberant R atio
EMA	E cological M omentary A ssessment
FAHL	F requency A verage H earing L oss
FIR	F inite I mpulse R esponse
FFT	F ast F ourier T ransform
HA	H earing A id
HL	H earing L oss
HATS	H ead A nd T orso S imulator
HOA	H igh O rders A mbisonics
IIR	I nfinite I mpulse R esponse
IR	I mpulse R esponse
NH	N ormal H earing
NMPS	N ormalised M odulation P ower S pectrum
RF	R adio F requencies
RIR	R oom I mpulse R esponse
RMS	R oot M ean S quare
RST	R ealistic S peech T est
RT	R everberation T ime
SHF	S pherical H armonic F unction
SI	S peech I ntelligibility
SNR	S ignal-to- N oise R atio
SPL	S ound P ressure L evel
SRT	S peech R eception T hreshold
SSQ	S peech S patial Q ualities
STI	S peech T ransmission I ndex
VSE	V irtual S ound E nvironment

To all the people who feel isolated due to hearing difficulties

Chapter 1

Introduction

1.1 Background and motivation

Despite the extremely complex acoustic characteristics of noisy and reverberant sound fields, most people have the ability to identify different simultaneous sounds and direct their attention towards them with merely the one-dimensional representation provided by each eardrum. This extraordinary ability enables people with normal hearing (NH) to maintain successful verbal communication in the presence of multiple competing simultaneous sound sources, a situation commonly referred to as the cocktail party problem (Cherry, 1953). However, the situation is completely different for cochlear implant (CI) users, for whom speech understanding in situations with multiple speakers and reverberant environments is exceptionally challenging. The lower spectral resolution observed in CI users, along with the limited access to the temporal fine structure of the signal and their impaired ability to localize sound sources are among the main reasons why CI recipients struggle in adverse acoustic conditions. This is in contrast to their high speech intelligibility (SI) performance in quiet conditions, which in some cases has been shown to be comparable to normal hearing listeners (Wilson and Dorman, 2007). Consequently, current efforts to improve CI technology are focused on more challenging tasks such as speech recognition in noise, localization (Zeng et al., 2008; Van Hoesel and Tyler, 2003) and the effect of reverberation (Kressner, Westermann, and Buchholz, 2018; Kokkinakis and Loizou, 2011), among others.

One of the challenges of evaluating speech recognition in a laboratory environment concerns the complexity of both the setup and the methods required to run highly realistic tests as well as the limited knowledge that is available on how to design such tests. These are likely the reasons why, in most cases, SI tests are conducted with oversimplified methods whereby overly articulated anechoic target speech is presented from a single loudspeaker while two or three loudspeakers located elsewhere reproduce babble noise. Clearly, a setup like this cannot reproduce the complexity of real-world sound fields where dynamic sound sources of different sorts come from random locations at random times while interacting with the effect of the room. Moreover, SI tests are often adaptive, whereby either the level of the target speech signal or that of the noise is altered to reach the signal-to-noise ratio (SNR) at

which a certain level of intelligibility is achieved. While these procedures have been shown to optimise test sensitivity, they are not necessarily representative of what people experience in the real world.

Alternative methodologies aimed to provide insights about an individual's real-life experience include retrospective questionnaires (e.g., Gatehouse and Noble, 2004; Cox and Alexander, 1995) or field-studies that apply data logging, usually conducted by means of mobile devices (Galvez et al., 2012). These techniques can present results with high ecological validity but provide very poor control over the stimuli. Often, retrospective questionnaire items are of the form "You are in [specific situation]. Can you follow the conversation?". These types of questions are often not easy to answer because many variables are open to interpretation (e.g., the distance, the room, the noise level, the number of competing talkers or the person you are talking to). Field studies using real-time questionnaires combined with data logging, which are often referred to as ecological momentary assessment (Galvez et al., 2012), improve the association between the stimuli and the responses, but are time consuming to run, rely on a high level of cooperation by the subjects, and have difficulties to ensure that the subjects actually find the acoustic environments of interest. Moreover, a detailed analysis of the characteristics of the encountered speech and noise signals is limited, as government regulations often prohibit recording of the actual sound signals, and the analysis is therefore limited to algorithms blindly estimating basic signal features from the noisy speech mixtures.

Understanding the difficulties faced by CI users in their daily lives without compromising control over the stimuli is a complex endeavour. The approach adopted in the present study is based on the principle *bring the real-world into the laboratory*. As detailed later, this has been accomplished in multiple ways through both target and noise signal manipulations which, ultimately, have enabled (1) a faithful reproduction of the reverberation of real rooms, (2) the use of speech materials that were extracted from natural, unscripted conversations, (3) a three-dimensional reproduction of real-world acoustic environments and (4) an evaluation of SI under realistic SNRs. Nevertheless, applying the most achievable level of realism to test paradigms may in some cases be counter-productive, as the number of uncontrolled variables may become too large for researchers to be able to disentangle their own individual effects. As will be detailed later, the present study has progressively built up in complexity where some realism was initially traded off in benefit of a better control, and the opposite occurred towards the end of the study.

The advantages of conducting tests inside the laboratory under realistic conditions are manifold. First, it makes it possible to have a good understanding of the difficulties faced by CI recipients in the real-world. This enables researchers to focus on signal processing techniques specifically devised for speech enhancement in adverse situations that may occur in real life. Second, it eases the interpretability of laboratory-based SI outcomes in terms of real-life performance. This can be particularly helpful in the development process of new speech processing algorithms

where frequent testing is required, and in the judgement of new research findings in terms of their relevance in the real-world. Third, it allows full replicability of the tests, which is of great importance to ensure that all subjects (and/or same subject in longitudinal studies) and the hearing devices are tested under the same conditions, regardless of the complexity of the sound field. Fourth, it allows direct access to the acoustic signals that arrive at the listener's ears or hearing devices during testing for a detailed signal analysis, which may even allow access to the target signal in isolation of the noise or reverberation. This in turn can help to understand the involved auditory processes or to systematically evaluate and optimise the benefit provided by a hearing device.

However, because the ambition is to ensure that all the acoustic features of the sound environment are accurately reproduced in the laboratory, such an approach is not free of difficulties. These include, but are not limited to, ensuring a realistic reproduction of the spatial and frequency characteristics of the sound field, the sound pressure level, the level of the target speaker based on their distance, the reverberation characteristics of target and noise sources and the vocal effort of the target speaker. The relevance of accounting for each and every feature of real acoustic environments is explored further below.

1.1.1 Unrealistic reverberation

The temporal smearing effect of reverberation flattens the envelope of speech signals. Because SI with CIs relies heavily on the envelope of the speech signals, the detrimental effect of reverberation on SI is higher for CIs than for people with NH (Xia et al., 2018; Nabelek and Pickett, 1974; Kressner, Westermann, and Buchholz, 2018). However, the conditions under which most studies have evaluated the effect of reverberation on SI are not representative of what people may encounter in the real world. Outside the laboratory, reverberant rooms can be found in all sorts of contexts, from toilets to cathedrals. However, the concept of "reverberant space" is most commonly associated with large rooms like cathedrals or concert halls, where the reverberation tail is quite long. Most likely, a clinician asking a client about their listening experience in reverberant spaces would be interpreted as referring to spaces with long reverberation tails. While the length of the reverberation tail of a room can uniquely be described with the reverberation time (RT), the effect of reverberation on SI certainly cannot. This primarily occurs because the RT provides no information about the direct sound component, whose level, relative to the reverberant field, is relevant to SI (e.g., Hersbach et al., 2015). The wrong assumption that SI can uniquely be explained by the RT has been observed in a large number of studies concerned with CIs and it likely arises from the use of unrealistic reverberation conditions (Hu and Kokkinakis, 2014; Kokkinakis, Hazrati, and Loizou, 2011; Kokkinakis and Loizou, 2011). One of the implications of this assumption is the misconception that laboratory-based SI outcomes are translatable to the real world by means of the RT, which can be particularly

misleading in a variety of contexts, ranging from the definition of speech processor requirements to customer counselling.

1.1.2 Unrealistic sound environments

The noisy environments in which humans communicate differ not only in regards to sound pressure level, but also across temporal, spectral, and spatial characteristics. All of these attributes may be of significance when it comes to speech understanding. For example, the background noise in a shopping centre is often loud, diffuse, steady, unintelligible and provides increased power at low frequencies. The difficulty of speech communication in such space likely diverges from that experienced in a small café, where communication flow is constantly disrupted by impulsive sounds like cutlery and intelligible conversations of other groups of people located nearby that may prove distracting or even annoying.

Research studies conducting speech-in-noise tests with CI users usually employ a single loudspeaker reproducing target speech while two (e.g., Rana et al., 2017) or three (e.g., Mauger et al., 2014) loudspeakers that are located elsewhere reproduce the noise signals, which typically consist of a one person's discourse or babble noise comprising several talkers. Although tests conducted with these noise materials and layouts certainly provide invaluable information about the person's abilities to deal with noise, they are very specific and artificial.

Further, generalising the results obtained with these noise materials to the real world would entail the risk of assuming that the only attribute of the noise signal that varies across environments is the sound pressure level. This misconception may give rise to particularly misleading conclusions in cases where relevant and unacknowledged noise attributes are much too different between real life and the laboratory. One particular aspect that is commonly overlooked is the effect of reverberation on the noise signals, which may have a great impact on the extent to which noise signals like babble noise impair intelligibility in CI users. Hence, the SNR at which the listener achieves a certain SI score may differ across reverberation conditions. Another example involves the SNR improvement provided by a beamformer, which may mistakenly be assumed to be translatable to any noisy environment, regardless of its spatial characteristics. The tests where this benefit is estimated are often conducted under very specific conditions that may represent a rough approximation of certain realistic noise environments, but certainly not all of them. In fact, several studies have reported that laboratory-based assessment of the benefit obtained with directional microphones does not reflect the benefit observed in real life (Walden et al., 2000; Cord et al., 2002; Cord et al., 2004).

1.1.3 Unrealistic speech material

It is often the case that CI users who participate in research studies do not see a connection between the speech employed in the tests and the speech that they encounter

in their daily lives. This is not a surprise considering the significant differences between the two cases. Sentences used in SI tests are scripted, well-formed and self-contained, and are read by a trained speaker with a clear voice, well articulated, and at a slow pace. This greatly diverges from speech in the real-world, which is typically quite rapid, consists of large variations in syntactic constructions, carries repeated and redundant information, and includes phonetic reductions and deletions (for an in-depth review, see Beechey, 2019).

One aspect that is also typically overlooked when using speech materials is how the speech level interacts with the noise level. For example, when two people are conversing in background noise, the interlocutors raise their voice in order to “talk above the noise”. This effect, known as the Lombard effect (Lombard, 1911), is not just characterised by an increase in sound pressure level, but also other acoustic properties such as changes in fundamental frequency, vowel duration and spectral tilt, to name a few (Lu and Cooke, 2008).

1.1.4 Unrealistic SNRs

It is typically the case in real-world settings that the signal and noise levels are not entirely independent. As mentioned above, people tend to “talk above the noise” when communicating in background noise (see Smeds, Wolters, and Rung, 2015; Weisser and Buchholz, 2019 for a comprehensive review) or move closer to their communication partner. Whereas both strategies lead to an increase in the effective SNR, increasing the vocal effort level also increases the overall background noise for other people, who may then raise their voices again to adjust to the new noise level. The dependency between target and noise levels observed in the real world is not incorporated in current laboratory-based assessment of SI, which is often centred around testing Speech Reception Thresholds (SRT; e.g., Keidser et al., 2013). While an SRT ensures that performance is measured at the most sensitive point of the psychometric function (i.e., the 50% point) and performance measures will not reach floor or ceiling, it typically results in a highly unrealistic SNR (Smeds, Wolters, and Rung, 2015), and refers to a condition that most people would not be able to communicate in at all or for a very long time.

1.1.5 Unrealistic listening tasks

According to Kiessling et al. (2003), verbal communication in everyday situations involves four distinct levels of hearing-related functioning: *hearing*, *listening*, *comprehension* and *communication*. *Hearing* is mainly characterised for being passive provided that no cognitive resources are necessary for sounds, as a percept, to exist. *Listening* is active and reflects speech understanding. *Comprehension* is also active and refers to the ability to comprehend speech. The last level, referred to as *communication*, is characterised by the fact that verbal communications, for example conversations, are interactive. Any given level depends on the levels located below it.

For example, comprehension depends on the ability to hear and the ability to understand what is being said. In contrast, a given level does not depend on the levels above it. For example, hearing does not imply listening, listening does not imply comprehending, etc.

Speech-in-noise tests are based on word recall, which falls somewhere between *listening* and *comprehension*, because, even though there is no actual need to comprehend the meaning of a sentence to recall the individual words, extracting the meaning can assist in using language redundancy to guess words not properly understood. Of course, this assumes that the provided sentences provide redundant information or context, which is often deliberately not the case (e.g., Hagerman, 1982). In any case, sentence tests that assess verbatim recall performance cannot guarantee comprehension nor do they include the interactive processes involved in real-life communication. They mainly assess low-level auditory function and the associated low-level deficits associated with a hearing loss.

1.2 Approach

This project represents a step forward in the process of enabling highly realistic listening tests inside the laboratory. As part of a process of continuous learning, this project evolved from rather specific research questions to more exploratory analyses. A large portion of this thesis aims at having a better understanding of the effect of realistic reverberation on SI. In this case, the concept of realistic SNRs and realistic speech material were traded off in benefit of a more systematic investigation in which the only difference across conditions was entirely the result of reverberation. Despite this constraint, realism was maximised by incorporating three dimensional representations of reverberant sound fields encountered in a wide variety of real rooms. This allowed conclusions on the main room acoustic parameters affecting SI in CI recipients, and gave rise to a first version of a simple room-based SI model, which was able to explain the data measured under a large number of reverberant conditions in quiet and noise. The remaining part of this thesis measured SI in CI users in a number of highly realistic conditions where the reproduction of acoustic environments was combined with realistic speech, including realistic reverberation, and presented at ecologically-valid SNRs. The SI outcomes measured in this new speech-in-noise paradigm were compared against SI outcomes of two other, less realistic speech-in-noise paradigms. Hence, the goal in this case was not only to report SI under more realistic conditions, but also to understand the differences in outcomes between a "standard" speech-in-noise test that is commonly applied in the laboratory and the new, more realistic, speech-in-noise tests, which included differences seen on the individual subject level.

In what follows, the different methods followed to achieve such levels of test realism are further explained. Note that, as part of a progressive increase of test paradigm complexity, the inclusion of realistic listening tasks was here not possible.

Future research should consider the possibility of including more realistic tasks such as comprehension (Best et al., 2016) or communication (Beechey, Buchholz, and Keidser, 2019) tasks.

1.2.1 Realistic reverberation and sound environments

Successful understanding of speech in adverse situations relies on a multitude of auditory cues that involve complex temporal, spectral and spatial auditory processes (Bronkhorst and Plomp, 2005), as well as on cognitive processes such as (selective and spatial) attention or short-term memory (Pichora-Fuller and Singh, 2006). To ensure that listeners taking part in laboratory-based experiments can exploit these cues as if they would in real life, researchers have already started employing 3D audio technologies to reproduce realistic acoustic/sound environments inside the laboratory. Three-dimensional audio techniques can be categorised into *binaural audio*, which relies on the use of headphones (e.g., Mueller et al., 2012; Rychtáriková et al., 2009), and *sound field reproduction* methods, which are based on multi-channel loudspeaker systems (e.g., (Oreinos and Buchholz, 2016; Favrot and Buchholz, 2010; Seeber, Kerber, and Hafter, 2010; Grimm, Ewert, and Hohmann, 2015)).

The approach adopted in the present study is a sound field reproduction method based on the concept of Higher Order Ambisonics (HOA). Higher Order Ambisonics is a tool that enables the codification of the spatial characteristics of a sound field, which can then be reproduced by means of an array of loudspeakers. The HOA process entails decomposing a recorded (real) sound field into a set of harmonic functions (i.e., the encoding stage) and thereafter finding the loudspeaker gains (i.e., the decoding stage) such that the directivity of the resulting sound field (namely, its ambisonic components) matches that of the original sound field as accurately as possible (see Appendix A for more details). Higher Order Ambisonics and variations thereof have already been implemented and validated with special attention to hearing research (Oreinos and Buchholz, 2016; Favrot and Buchholz, 2010; Grimm, Ewert, and Hohmann, 2015) although alternative multichannel-based techniques exist as well (e.g. (Seeber, Kerber, and Hafter, 2010)).

One of the benefits of HOA with respect to alternative sound field reproduction methods is that the encoding stage is completely independent from the decoding layout, which enables the exchange of HOA-encoded sound environments across institutions that have different loudspeaker configurations for playback. In fact, this feature of HOA made it possible to release the Ambisonic Recordings of Typical Environments (ARTE) database, a set of 13 HOA recordings of realistic environments that is available online (Weisser et al., 2019). In short, the ARTE database enables the reproduction of previously recorded typical noisy environments (e.g., cafe, living room, food court, office, etc.) over two or three-dimensional arrays of loudspeakers or over headphones.

Similar to the ARTE database, the present study employed an array of 62 microphones flush-mounted on the surface of a rigid sphere to record real acoustic scenes,

encoded the 62 channels into 31 HOA components and decoded the HOA-encoded signals into a spherical array of 41 loudspeakers. Speech intelligibility tests were conducted with subjects sitting in the centre of the 41 loudspeaker array located in the anechoic chamber of the Australian Hearing Hub (Macquarie University, Australia).

Higher Order Ambisonics was also applied for the auralization of the target speech material by convolving it with HOA-decoded room impulse responses (RIRs). The first step to obtain these RIRs was to place the microphone array at the intended listener position within each room when there was no noise present. A loudspeaker was placed at the intended talker position within the room and used to excite the room with a known signal (i.e., a logarithmic sweep) in such a way that a 62-channel RIR could be measured. This RIR was then HOA encoded and subsequently decoded giving rise to a 41-channel RIR that was readily available for convolution with anechoic target speech signals. Because the RIRs were obtained in real physical rooms, it was ensured that the time, frequency and spatial characteristics of the reverberant sound fields corresponded to the real case. Moreover, because the loudspeaker array was located in an anechoic chamber, the reverberation characteristics of each simulated room were largely preserved. In both noise recordings and RIR measurements, several equalisation and calibration procedures enabled a highly accurate reproduction of the original (real) sound fields at their natural levels.

Importantly, artificial noise scenes were also employed in the present study, including scenes that are representative for the rather artificial listening tests that are commonly applied in the laboratory. In particular, dialogues between two people recorded in anechoic conditions were convolved with HOA-decoded RIRs and used as background noise signals in the study presented in Chapter 2. This is an example where some level of realism was traded off against stimulus control, as in this case the requirements of the test constrained the different noise conditions to differ only in their reverberation characteristics. However, in Chapter 4, an experiment is reported in which this constraint was removed to achieve the highest possible level of realism that could be achieved using the available technologies.

1.2.2 Realistic speech material

As discussed in Sec. 1.1.3, speech materials typically used in clinical and laboratory experiments comprise contrived sentences (e.g., "The clown had a funny face"), read by a trained professional in a quiet environment. In order to test CI recipients under more realistic conditions, in Chapter 4 an experiment is reported that applied a newly developed, more realistic speech test in the highly realistic acoustic/sound environments described in Sec. 1.2.1.

The test incorporates two levels of realism. The first is that the sentence material was extracted from natural conversations between a pair of native Australian-English speakers. As there were no restrictions on how the pair were to converse,

sentences were occasionally slow and mumbled, and other times quick with repetitions. The sentences were also not always “well-formed” with the typical subject-verb-object structure of most speech tests (e.g., “that’s interesting isn’t it”). The second level of realism comes from the manipulation of vocal level. The pair each wore highly open headphones when conversing that played one of three background noises from the ARTE database (community centre, café, food court). As each environment has a different sound pressure level, the talkers modified their vocal level in order to converse at a comfortable level (see Sec. 1.2.3 below). Thereby, the community centre elicited a “normal”, the café a “moderate” and the food court a “raised” vocal level, which was in line with the vocal effort levels reported in the standard ANSI S3.5 (1997).

This newly developed sentence test has previously been validated with 32 young normal-hearing individuals, who achieved near 100% SI in quiet for each sentence with an average psychometric function in speech-shaped noise exhibiting an overall slope of 16%/dBSNR (Kelly M. Miles et al., 2019, in preparation).

1.2.3 Realistic SNRs

Weisser and Buchholz (2019) measured the sound pressure levels of natural conversations between two people while being presented with 13 different realistic noisy environments via highly open headphones. In agreement with the literature on the Lombard effect (Lombard, 1911), they derived speech levels, and thus SNRs, that were strongly dependent on the noise level. In their paper, the authors provide an estimate of the SNR not only as a function of noise level, but also as a function of the talker-to-listener distance (i.e., the closer the higher the SNR) and whether the talker is a female or a male, as they were shown to provide slightly different levels. These relationships between signal and noise levels were applied to select ecologically valid speech levels (or SNRs) for the different realistic acoustic environments used in the experiment reported in Chapter 4.

1.3 Aims

The overarching goal of this study was to investigate the effect of reverberation and noise considering realistic acoustic scenes. Depending on the question at hand, different levels of realism were applied. The first aim of the study was the evaluation of the effect of realistic reverberation on SI with CIs in both quiet and noisy conditions. This study was motivated by the conflicting results obtained across studies concerning the effect of reverberation in quiet conditions. Because the question was very specific, the test paradigm was designed so that the conditions included in the study differed only on their reverberation characteristics. The second aim was to evaluate the accuracy of a first version of a simple room-based SI model as a predictor of SI with CIs, as well as to evaluate the main room acoustic factors affecting SI in CIs in both quiet and noise. The third aim was to assess the effect of test realism on SI

outcomes, which was conducted by comparing SI outcomes obtained under three different levels of test realism. The fourth aim was to evaluate SI with CIs in highly realistic conditions and to assess how much benefit recipients obtained by the use of two devices instead of only one.

1.4 Outline

Chapter 2 presents a study devoted to evaluating the effect of reverberation on SI with CI users tested unilaterally under both quiet and noisy conditions. Chapter 3 is based on the same dataset as Chapter 2 and presents a first version of a room-based SI model that is especially suitable for predicting the effect of reverberation on SI in quiet. Chapter 4 presents a study whose goal is twofold. First, to evaluate the effect of test realism of SI outcomes. Second, to evaluate SI of bilateral CI users under highly realistic conditions as well as the benefit in terms of SI obtained by using two devices as opposed to only one.

Chapter 2

Effect of noise and reverberation on speech intelligibility for cochlear implant recipients in realistic sound environments

Abstract

Previous studies have suggested a strong effect of reverberation on speech intelligibility (SI) in cochlear implant (CI) recipients. In most studies, different reverberation conditions were obtained by altering the acoustic absorption of a single room, thereby obtaining different reverberation times (RT). In these cases, higher RTs imply higher-pressure reverberant fields, a condition that does not necessarily occur in real-life. In addition, studies that have investigated the combined effects of reverberation and noise on SI have not examined the effect of reverberation on the temporal fluctuations of the noise. The present study investigates the realistic, reverberant conditions in which CI recipients have difficulties understanding speech in quiet and noise. Percent correct sentence recall scores were measured in 12 unilateral CI recipients both in quiet and in noise using a 3D loudspeaker array in an anechoic chamber. Target speech was convolved with room impulse responses (RIRs) recorded at three talker-to-listener distances in five physical rooms with distinct RTs and presented at 60 dB-SPL. Noise consisted of four two-talker dialogues convolved with RIRs measured at four fixed positions around the listener. Results in quiet suggest that a significant drop in SI occurs mainly at large talker-to-listener distances, and small reverberant rooms affect SI the most, which highlights the importance of the level of the direct sound relative to that of reverberation. In noise, results show that the most detrimental type of noise is anechoic, as it is the most modulated. A comparison between rooms in terms of SI scores, as well as self-reported listening effort, in both quiet and noise suggests that CI users can handle reverberation rather well at short distances in rooms with large volume or small rooms with some reverberation.

2.1 Introduction

In most everyday listening environments sound signals experience multiple diffractions and reflections bouncing off walls, ceilings and floors, which, as a whole, are referred to as reverberation. Reverberation can be seen as a set of *uncountable*, delayed, frequency-dependent replicas of the direct signal that bounce repeatedly within the enclosure until they dissipate. As such, the reverberant field differs from the direct field in time, frequency and space, which has multiple implications in terms of distance perception (Bronkhorst and Houtgast, 1999), lateralization (Hartmann, 1983) and speech intelligibility (Nabelek and Pickett, 1974). In a reverberant speech signal, formant transitions are flattened, the envelope is smoothed and, because absorption is less effective at low frequencies, increased upward spread of masking may occur due to reverberation (Greenberg et al., 2004, pp. 269–275). The main effects of reverberation on speech intelligibility (SI) are often broken down into overlap-masking and self-masking effects (Bolt, 1949). Overlap-masking is a phenomenon that occurs when the energy of a precedent phoneme masks a subsequent one. Self-masking occurs when the energy is smeared within a phoneme, which flattens the transition between formants.

While speech understanding in quiet reverberant spaces is rarely compromised in people with normal hearing (Nabelek and Pickett, 1974), the negative effect of reverberation on cochlear implant (CI) recipients can be significant (Kressner, Westermann, and Buchholz, 2018). Normal hearing (NH) listeners apply a detailed temporal and spectral analysis of the incoming signals to understand speech in reverberation (and noise), which is further assisted by a number of additional (monaural) auditory cues, such as fundamental frequency cues or temporal fine structure cues (Darwin and Hukin, 2000). In CI recipients, the temporal and spectral resolution is highly reduced and most of the additional monaural cues are not available due to the inability of current CI technology to convey temporal fine structure information. Due to the latter, the signal envelope is mainly encoded in CIs, which is strongly distorted by the temporal smoothing introduced by reverberation (Houtgast, Steeneken, and Plomp, 1980). Given these limitations, it is expected that SI performance of CI recipients degrades more quickly with increasing reverberation than for NH listeners.

The negative impact of reverberation on SI with CI recipients has already been evaluated in previous studies using a variety of methods. In some studies, SI in reverberation has been evaluated by convolving anechoic target speech with simulated Room Impulse Responses (RIRs) to test either simulated CI recipients, i.e., by using vocoded signals with NH subjects (Desmond, Collins, and Throckmorton, 2014; Helms Tillery, Brown, and Bacon, 2012; Whitmal and Poissant, 2009; Poissant, Whitmal, and Freyman, 2006), or with actual CI recipients (Hazrati and Loizou, 2013; Hazrati, Lee, and Loizou, 2013). In some other cases, SI tests are conducted with CI recipients, and the speech material is obtained by convolving anechoic speech

with in-ear (Hu and Kokkinakis, 2014; Kokkinakis, Hazrati, and Loizou, 2011; Kokkinakis and Loizou, 2011) or omnidirectional (Hazrati and Loizou, 2012) RIRs that were recorded in a real room. All these studies obtained different RTs by varying the absorption of a simulated or real room, and applied non-individualized stimuli that were either presented via headphones (to NH listeners) or the direct audio input of the CIs. Kressner, Westermann, and Buchholz (2018) is the only study that systematically investigated the effect of reverberation on speech intelligibility in CI recipients using a number of more realistic scenarios where individual spatial cues were provided. Interestingly, they found that reverberation has a far weaker impact on SI in CI recipients than previously reported.

In noise, it is well known that for satisfactory speech intelligibility, CI recipients generally need higher signal-to-noise ratios (SNRs) than NH listeners. In the presence of fluctuating noise, the difference between NH and CI is even higher (Fu and Nogaki, 2005). Whereas NH listeners can take advantage of temporal gaps present in fluctuating noises that allow them to improve SI, known as masking release, glimpsing, or dip listening (Bronkhorst, 2000; Cooke, 2006; Festen and Plomp, 1990), CI recipients present an exceptional sensitivity to modulated noises (Fu and Nogaki, 2005; Nelson et al., 2003; Qin and Oxenham, 2003). This may be explained again by the fact that CI recipients do not have access to the signal's fine structure, which makes it impossible for them to identify the changes in temporal fine structure in the dips of a noise signal (Hopkins and Moore, 2009) and strongly degrades the ability to segregate the target speech from the noise. Given that reverberation distorts the envelope of the target speech as well as increases the overall level of the noise, it is expected that reverberation is particularly detrimental to CI listeners in noisy conditions. However, the observation that reverberation smooths the envelope of modulated noises may aid SI in rooms and partially compensate the decrease in SNR. Moreover, CI users may be able to take advantage of the early reflections that may effectively increase the power of the speech signal (e.g., Kressner, Westermann, and Buchholz, 2018).

The impact of noise on the speech intelligibility performance in CI recipients has been widely studied in research laboratories around the world, and is routinely assessed in audiological clinics. However, very few studies have investigated the combined effect of reverberation and noise, and the existing studies either applied reverberation only to the target speech but not to the noise (Hazrati and Loizou, 2012), or used vocoding to simulate CI recipients with NH listeners (Whitmal and Poissant, 2009; Poissant, Whitmal, and Freyman, 2006). Hence, very little is known about the combined effect of reverberation and noise on speech intelligibility performance in actual CI recipients.

The goal of the present study is to systematically evaluate the ability of unilateral CI recipients to understand speech in both quiet and noise under a number of reverberant conditions. In order to improve the ecological validity of the outcomes over previous studies, subjects wore a real-time speech processor that mimicked their own

processor and were presented with realistic three-dimensional (3D) sound fields that were created from real acoustic scenes. RIRs were recorded with a 3D microphone array in a variety of rooms and at multiple talker-to-listener distances and convolved with anechoic speech material. The reverberant speech was then reproduced with a 3D loudspeaker array inside an anechoic chamber using the higher-order Ambisonics method (Oreinos, 2015b). The background noise was realized in the same way and consisted of four pairs of talkers who had one-on-one conversations and were added to the target speech for each of the rooms individually. Using this method, the subjects were given the impression of being in the actual room, and they were able to utilize their own individual spatial cues including head movements. Because the acoustic scenes were obtained from a range of real rooms, they represent reverberant conditions that CI recipients are likely to experience in their daily lives. Using modulated noise (i.e., 4-talker babble) allows the study of the effect of the room (i.e., reverberation) on the modulation depth of the noise in terms of SI.

2.2 Methods

2.2.1 Participants

Twelve postlingually deafened CI recipients participated in this study who had at least 12 months experience with their devices. All participants were tested unilaterally. Eight of the 12 participants were bilaterally implanted, and were tested with their preferred ear. Two participants were bimodal CI recipients (S9 and S12) and wore a Hearing Aid (HA) on the contralateral ear. Their four-frequency average hearing loss (4FAHL) was 96 dB HL and 61 dB HL, and their best frequency band above 250Hz was 70 dB HL and 40 dB HL, respectively. These two participants were tested with their HAs removed and no earplugs were used. The remaining two participants were unilateral CI recipients. One of them was completely deaf in the non-implanted ear (S6), and the other one (S8) had a 4FAHL of 85dB HL with 20dB HL and 30dB HL at 250Hz and 500Hz, respectively. E-A-R™ Classic™ Platinum Earplugs were used during the test of S8, which provided an attenuation of at least 20 dB.

The testing was divided into two visits of at most 2 hours each. Participants were paid in appreciation of their participation. All participants were users of Cochlear devices, used CP810 or more recent speech processors, and were users of the Advanced Combination Encoder (ACE™) speech processing strategy.

2.2.2 Stimuli

2.2.2.1 Speech material

The material for the target speech is known as the “BKB-like sentences” and was developed by the Cooperative Research Centre for Cochlear Implant and Hearing

TABLE 2.1: Relevant biographic data of the participants

ID	Gender	Mode	Age	Tested ear	Implant in tested ear	Age at time of implant	Cause of hearing loss
1	F	Bilateral	61	L	CI422	56	Acquired
2	F	Bilateral	73	L	CI422	69	Acquired
3	M	Bilateral	60	R	CI522	59	Acquired
4	F	Bilateral	42	R	CI24R	26	Acquired
5	M	Bilateral	54	R	CI522	53	Congenital
6	F	Unilateral	71	R	CI24M	53	Acquired
7	F	Bilateral	58	R	CI24RE	51	Acquired
8	F	Unilateral	64	L	CI512	57	Acquired
9	F	Bimodal	68	R	CI24RE	62	Acquired
10	F	Bilateral	59	R	CI24RE	53	Acquired
11	F	Bilateral	55	L	CI24RE	43	Acquired
12	M	Bimodal	76	R	CI422	71	Acquired

Aid Innovation (CRC HEAR) in a similar manner as the Bamford–Kowal–Bench sentences (Bench, Kowal, and Bamford, 1979). The corpus consists of 80 lists of 16 sentences recorded by an Australian female speaker at a sampling frequency of 44.1 kHz. Sentences comprise up to six words or eight syllables and contain vocabulary that is familiar to a five-year-old.

The speech material for the background noise consisted of four two-talker dialogues extracted from a series of IELTS™ passages. The dialogues were reproduced by native Australian English talkers in a large anechoic chamber and recorded at a sampling frequency of 44.1 kHz. Within each dialog there was very little talker overlap, so that the combined masker contained basically four concurrent speech streams.

2.2.2.2 Sound reproduction

Speech intelligibility was tested with subjects sitting in the center of a spherical array of 41 loudspeakers located in the anechoic chamber of the Australian Hearing Hub (Macquarie University, Australia). For every target and interferer position, corresponding anechoic speech materials were convolved with three-dimensional 41-channel RIRs. The first step to obtain these loudspeaker-specific RIRs was to record a microphone-specific RIR with an array of 62 microphones flush-mounted on the surface of a rigid sphere. The microphone array was placed at the intended listener position inside each room at 1.3 meters above the floor and as far as possible from the walls. A loudspeaker (Tannoy V8) was placed at the same height to excite the room with a logarithmic sweep, and the transfer functions between the recorded and the reproduced sweeps were calculated and transformed into the time domain using an inverse Fourier transform to obtain a set of 62 RIRs (one for each

microphone). This procedure was repeated for eleven loudspeaker locations in each room, corresponding to the 3 target and 8 interferer locations used in the SI tests as illustrated in Fig. 2.1. The 62-channel RIRs were encoded into the Mixed Order Ambisonics format ($M2D = 7$ and $M3D = 4$, Favrot et al., 2011) and subsequently decoded into 41 reproduction channels, each of them corresponding to a single playback loudspeaker. Thereafter, the recording noise floor level (arising from various noise sources such as the microphone) was identified and used to truncate the RIRs to their noise-free length, which was conducted in third octave bands and for each loudspeaker channel independently. The direct sound component of each decoded RIR was extracted from the simulation of the sound pressure at the center of the array by applying a one-sided Hanning window with a frequency-dependent duration of $D = \max(0.003, 2/f)$ seconds, and equalized. Since interferers consisted of pairs of talkers facing each other, they were rotated 90° relative to the listener (see Fig. 2.1). To accommodate for the frequency-dependent directivity of a talker, which is slightly different from the one of the loudspeaker used during the RIR recording, the direct sound component of the interferers was filtered such that it had a frequency response equal to a talker at 90° (Chu and Warnock, 2002). The filtering did not significantly alter the broadband energy of the direct sound component. The extracted direct sound component for each source location, including targets and interferers, was then assigned to a single loudspeaker in the playback array and recombined with the rest of the decoded RIRs. The processing of the direct sound component effectively extended the sweet-spot of the applied Mixed-Order-Ambisonics approach (Favrot et al., 2011), which enabled participants to move their head almost freely during the tests. Minimum-phase FIR filters were applied to the resulting 41-channel RIRs to equalize the individual loudspeakers of the playback array prior to convolution with the anechoic speech materials.

The same layout was used to record the RIRs in five acoustically distinct rooms. Table 2.2 shows their relevant acoustic properties measured at the listener position including the direct-sound-to-reverberation energy ratio (DRR) as well as the clarity ($C50$), measured as the energy ratio of the early reflections arriving within the first 50ms after the direct sound and the rest of the RIR (e.g. Bradley, 2002). The first room is a rather small (164m^3) lecture room (LR) acoustically treated with carpet, acoustic absorbers (two walls and ceiling) and a heavy acoustic curtain covering one of the largest walls. The second room is a workplace kitchen (WPK) area of 164 m^3 that, apart from the carpeted floor, is not acoustically treated. The walls and the ceiling of WPK are a combination of glass, plasterboard and rendered brick wall. It is common to hear complaints of the acoustic conditioning of WPK when social gatherings take place in it. The third room is a completely empty, large (446 m^3), L-shaped open-plan office (OPO). The floor is carpeted but the walls and ceiling are a combination of plasterboard, rendered brick wall, and, in a smaller proportion, glass. The fourth room is an empty, small (134m^3) reflective room (SRR) that is not acoustically treated at all. The floor, the walls and the ceiling are a combination of rendered brick walls,

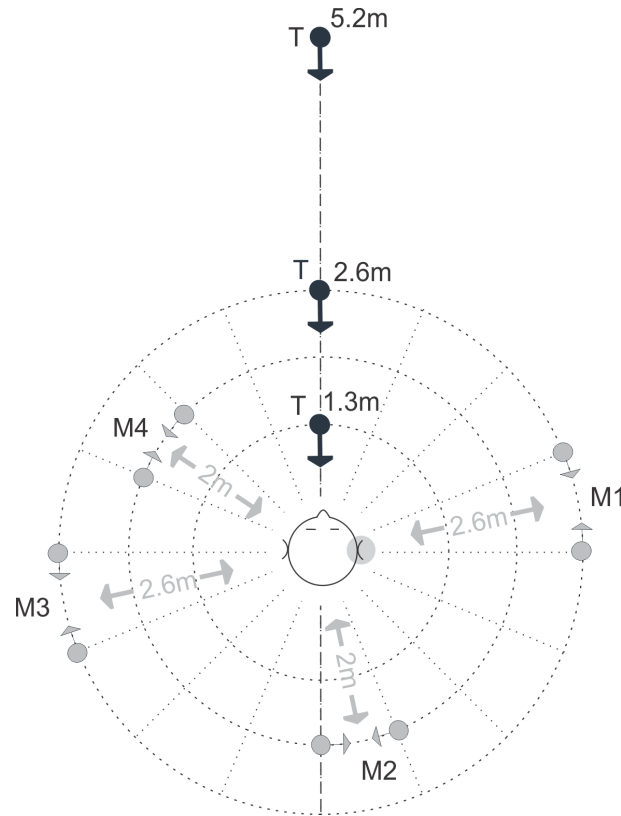


FIGURE 2.1: Layout of sound sources used during the experiments for listeners using the right ear. Target speech was presented in front of the listener at 1.3, 2.6 and 5.2 meters. Competing talkers consisted of four (M1 to M4) dialogues of two talkers facing each other and located at distances of 2.6m (M1 and M3) or 2m (M2 and M4) from the listener. The target speech in the car park (CP) was presented at an additional distance of 10.4 meters (not shown). For listeners using the left ear, source positions were mirrored around the front-back axis

plasterboard and linoleum. SRR is just a vestibule that leads into a main room; verbal communication in SRR is very challenging. The fifth room is an extremely large indoor car park (CP) with a very long RT, giving an acoustic perception very similar to that obtained in a church or even in a cathedral.

During the SI test, target and noise levels including reverberant energy were kept constant at the listener position, which was verified using an omnidirectional microphone located at the centre of the loudspeaker array. This was done regardless of the acoustic scene or target distance in order to reduce the confounding effect of audibility when target distance and room vary. The level of the target was always 60 dB SPL, irrespective of the participant. The level of the noise was determined for each subject individually by first measuring the Speech-Reception-Threshold, SRT50 (Keidser et al., 2013), using anechoic BKB-like sentences for the target speech, and the noise in the WPK from the main test. During the SRT measurement, anechoic targets were held fixed at 60 dB SPL and the WPK noise was adaptively varied until 50% correct SI was achieved. Each SRT test comprised up to 32 sentences, was scored per morpheme, had an initial SNR of 5 dB and employed various step sizes down to 1dB, as

explained in Keidser et al. (2013). The noise level used in the main SI tests was then reduced by 3 dB relative to the SRT50 level to reduce the likelihood of floor effects arising from the combined effects of noise and the reverberation of the target speech. Within a pilot test, using the WPK noise as this anchor point showed the highest test-retest sensitivity and minimised floor and ceiling effects across all conditions of the main experiment.

In addition to the five rooms, SI was also evaluated both in quiet and in noise for an anechoic condition (i.e. without a room). In the absence of RIR convolution, which would normally provide the appropriate distance cues, the closer proximity of interferers M2 and M4 (Figure 2.1) compared to M1 and M3, was simulated by increasing the relative amplitude of the closer interferers by a factor of 1.3, corresponding to the ratio of their distances (2.6/2). As described above and shown in Fig. 2.1, a filter emulating the frequency response of a talker that is rotated by 90° was also applied to each anechoic interferer.

Five additional conditions were included to help separating out the effects of target and interferer reverberation. These conditions consisted of anechoic target speech in the presence of reverberant interferers in the five different rooms. In total, there were seven additional conditions that were added to the 16 quiet conditions (4 rooms × 3 distances + 1 room × 4 distances) and their 16 noisy counterparts, leading to 39 test conditions. One BKB-like list (16 sentences) was used per condition.

2.2.3 Cochlear Implant Research Platform (CIRP)

In this study, the subjects' speech processor was replaced by a corresponding real-time emulation using the Cochlear Implant Research Platform (CIRP: Goorevich and Batty, 2005). The CIRP consists of (1) a BTE (Behind The Ear) sound processor shell including two microphones, (2) a dual-channel microphone preamplifier, (3) a PC specifically designed for real-time applications called Speedgoat™ xPC target computer, (4) a generic PC with Matlab® called host computer, and (5) a stimulus generator and radio frequency (RF) transmitter. Sound signals are received by the microphones of the speech processor, amplified and sent to the analog inputs of the xPC target computer. The xPC target computer processes the signal in real-time, in the same way as an actual speech processor would, and sends stimulation information to the signal generator and RF transmitter so that it can be sent to the implant in the listener. The CIRP used in this study, as well as additional hardware required to conduct safety tests, were provided by Cochlear Ltd. (Sydney, Australia). The xPC and the host computer were both located in the control room and connected to an isolation transformer. All the other devices (BTE, preamplifier and stimulus generator) were located in the anechoic chamber, and connected to a second isolation transformer.

The speech processor model is designed in the host computer using high-level visual language (Simulink®), with some functions written in Matlab® and C. The model is compiled into real-time code by means of Simulink Coder™ and a C/C++

TABLE 2.2: Acoustic properties of the five rooms used during the SI tests. The DRR and the C50 are given for each of the talker-to-listener distances. The parameters shown are the mean of the parameters obtained in octave bands between 125 Hz and 8 kHz. All parameters are obtained from the RIRs by conducting simulations of an omnidirectional microphone located at the center of the loudspeaker array (i.e., the location of the listener's head).

Room	RT (s)	Room volume (m ³)	Critical distance (m)	Distance (m)	DRR (dB)	C50 (dB)	ID
Lecture Room	0.46	164	2.78	1.3	6.9	15.4	1
				2.6	-0.6	11.5	2
				5.2	-4.6	8.6	3
Workplace kitchen	0.68	164	1.77	1.3	2.7	10.9	4
				2.6	-3.8	6.7	5
				5.2	-6.5	5.4	6
Open-plan office	0.96	446	2.49	1.3	5.2	13.9	7
				2.6	-1.5	10.8	8
				5.2	-4.5	8	9
Small reflective room	1.55	134	1.18	1.3	-0.3	3.8	10
				2.6	-6.2	1.6	11
				5.2	-9.8	0	12
Indoor car park	2.42	> 5700	3.34	1.3	7.2	11.5	13
				2.6	0.4	6.9	14
				5.2	-4	3.9	15
				10.4	-6.1	2	16

compiler. The code is transferred to the target computer using xPC TargetTM, and runs in the xPC (Simulink Real-TimeTM) operating system. To reduce the impact of potentially confounding noise reduction and signal conditioning in the speech processor, the following features were disabled: Microphone directivity, Automatic Dynamic Range Optimization (ADRO), SNR-NR, Spatial-NR, SCAN, WhisperTM and WNR (see Dawson, Mauger, and Hersbach, 2011, for details on the different technologies). Only the Automatic Sensitivity Control (ASC) was enabled, which prevents signals louder than 65 dB SPL from being clipped (Wolfe et al., 2015), and only the front microphone (omnidirectional response) was used.

In order to calibrate the CIRP signal levels, an internal 1 kHz pure tone generator emulating a 65 dB SPL pure tone was used as a reference. The output of the generator was extracted through an analog output of the xPC computer and its amplitude evaluated by digitizing the signal with an RME sound card connected to a PC with Matlab®. A similar procedure was followed using the CIRP speech processor suspended in the center of the loudspeaker array, which presented diffuse third-octave filtered noise centered at 1 kHz (ANSI S3.35, 2004) at 65 dB SPL. The sensitivity of the analog input of the xPC system was adjusted to match the output levels for the internal and external signals. The CIRP was fitted to the individual subjects by using the fitting parameters of their own devices, which were provided by their audiological clinic. The following parameters were included: T and C levels, number of Maxima, stimulation mode, stimulation rate, and volume and sensitivity settings. In addition to the CIRP functionality and safety tests conducted at Cochlear Ltd., subject-specific safety tests were carried out on the stimulation commands sent to the RF receiver to ensure that the fitting parameters were inserted and encoded correctly.

2.2.4 Procedures

After written consent was given by the subject, anonymous .cdx files were obtained to access the subject's fitting parameters. Before each subject's first session, the real time model was prepared by inserting the subject's fitting parameters (for the preferred ear in case of bilateral implantees). Listeners were seated at the center of the loudspeaker array. They were asked to remove their own device/s and to put the BTE shell and the RF transmitter on their (preferred) implanted ear. Subjects were asked to notify the experimenter in case the sound did not resemble that of their own device, which never happened.

Prior to the main experiment, it was necessary to measure the SRT50 in order to determine the noise level used in the noisy conditions. The SRT50 was measured twice using the methods described in Sec. 2.2.2.2, and the results were averaged. For each subject, the fixed noise level was then chosen to be 3 dB lower than the level that produced the SRT50.

In each of the 39 test conditions, subjects were tested with one list of 16 sentences. After each sentence was presented, listeners were asked to repeat as much as they understood of the target speech. The operator then scored the number of correctly

understood morphemes on a graphical user interface running on a computer outside of the anechoic chamber. Within a session, conditions were randomized and, within a condition, sentences were randomized.

Upon completion of each list (i.e., condition), subjects were instructed to fill out a brief questionnaire rating the scene. The questionnaire contained four questions: (1) How echoic (reverberant) did you find this space? (2) How much effort did listening to the speech take? (3) How distracting was the background noise? And (4) How loud was this scene? For each question, a continuous rating scale was provided ranging from 0 to 10. Integers were highlighted and the extremes were accompanied with a short description of their meaning, thus clarifying the direction of the response (e.g. 0: not echoic at all, 10: extremely echoic). This questionnaire was given and explained to the subjects in the very beginning of the experiment.

2.2.5 Statistical analysis

Proportion of morphemes correctly understood y was linearized to improve statistical validity with floor/ceiling effects by applying the transformation:

$$SI_{lin} = \ln \left(\frac{y + \epsilon}{1 - y + \epsilon} \right) \quad (2.1)$$

Where ϵ corresponds to the smallest non-zero value of $(1 - y)$ across all subjects and conditions, as suggested in Warton and Hui (2011), and SI_{lin} is the linearized speech intelligibility. A linear mixed-effects model with subject as a random intercept was then applied to the linearized SI scores to evaluate the different factors affecting speech intelligibility. In particular, effects of distance, room and noise were evaluated, as well as their interactions. In the model, the variable room was treated as a category, distance was treated as a continuous variable and noise was a categorical variable used to distinguish quiet from noisy conditions.

The model was fitted following the Bound Optimization BY Quadratic Approximation (bobyqa) algorithm, available in the *lmer* package in R. T-tests were conducted in R with the *emmeans* package and multiple comparisons were Tukey-corrected.

Regarding the analysis of the questionnaire, the fourth question, which asked about the loudness of the acoustic scene, was omitted because virtually all subjects reported the same loudness value regardless of the condition (overall mean = 3.54, standard deviation = 0.23). The remaining three questions were all treated as continuous. As for the analysis of the speech intelligibility scores, each questionnaire item was analyzed with a linear mixed-effects model where subjects were treated as a random effect. There was no need to linearize the questionnaire data. Due to a visual impairment and verbal communication difficulties, subject 2 could not fill out the questionnaire.

2.3 Results

As explained in Sec. 2.2.2.2, the SNRs employed during the experiment were calculated from the individuals' SRTs. The resulting individuals' SNRs were, from subject 1 to subject 12, 7 dB, 7.3 dB, 4 dB, 1.6 dB, 8.2 dB, 3.5 dB, 6 dB, 0.7 dB, 5.5 dB, 2.5 dB, 6 dB and 2.6 dB respectively. The results of both the SI tests and the brief questionnaire that was administered after each SI condition are described below.

2.3.1 Speech intelligibility data

Figure 2.2 shows individual and group mean SI scores for each condition. The top row shows scores in quiet and the bottom row those in noise. Each group of bars represents a room condition (including an anechoic room). Rooms are sorted from left to right according to the RT (see Table 2.2), with Anechoic (no room) on the left and car park on the right. Within each group individual bars show results for different talker-to-listener distances, increasing from left to right, with anechoic considered as a theoretical nearest distance. In quiet, the anechoic target "distance" is identical to the Anechoic (no room) condition at the extreme left of the figure, but is included for each room to facilitate comparisons.

In quiet anechoic conditions all subjects achieved SI scores between 90% and 100%. In contrast, variations between subjects in reverberation are much larger, spanning almost 90% difference (e.g., see car park at 10.4m or small reflective room at 5.2m). Despite the high subject variability, a number of general observations can be made. First, at short distances (e.g. 1.3m), SI in quiet is rarely compromised by reverberation. Apart from the small reflective room, all the rooms lead to mean SI scores higher than 90% at 1.3m. At 2.6m, the lowest mean SI among all these rooms is still 80%. Second, the SI scores in the small reverberant room in quiet are far below those obtained in any of the other rooms. At 1.3m distance, mean SI scores are already below 70%, and at 5.2m, mean scores are below 40%, with some subjects understanding virtually nothing. Unlike in any other room, only one subject (S8) presented scores above 80% at all distances. Third, in quiet, higher RTs do not necessarily lead to lower SI scores. For example, the RT of the small reflective room is 1.55s, whereas that of the car park is 2.42s, and yet SI is generally lower in the small reflective room.

In noise (Figure 2.2, bottom), apart from the monotonic decay of SI over distance already observed in quiet, the order of rooms according to SI scores differs between quiet and noise. For example, while the workplace kitchen leads to rather high scores in quiet, scores in noise are the second worst. Likewise, while the SI scores achieved in quiet in the lecture room and the car park become increasingly different with distance, in noise they are comparable, indicating that the lecture rooms has a relatively more significant effect of noise. The different effects of noise on SI are most clearly observed in the mean SI score for anechoic target speech in anechoic noise (first bar of the lower plot in Fig. 2.2), which is much lower than for anechoic target speech in reverberant noise. These differences between quiet and noisy conditions, together

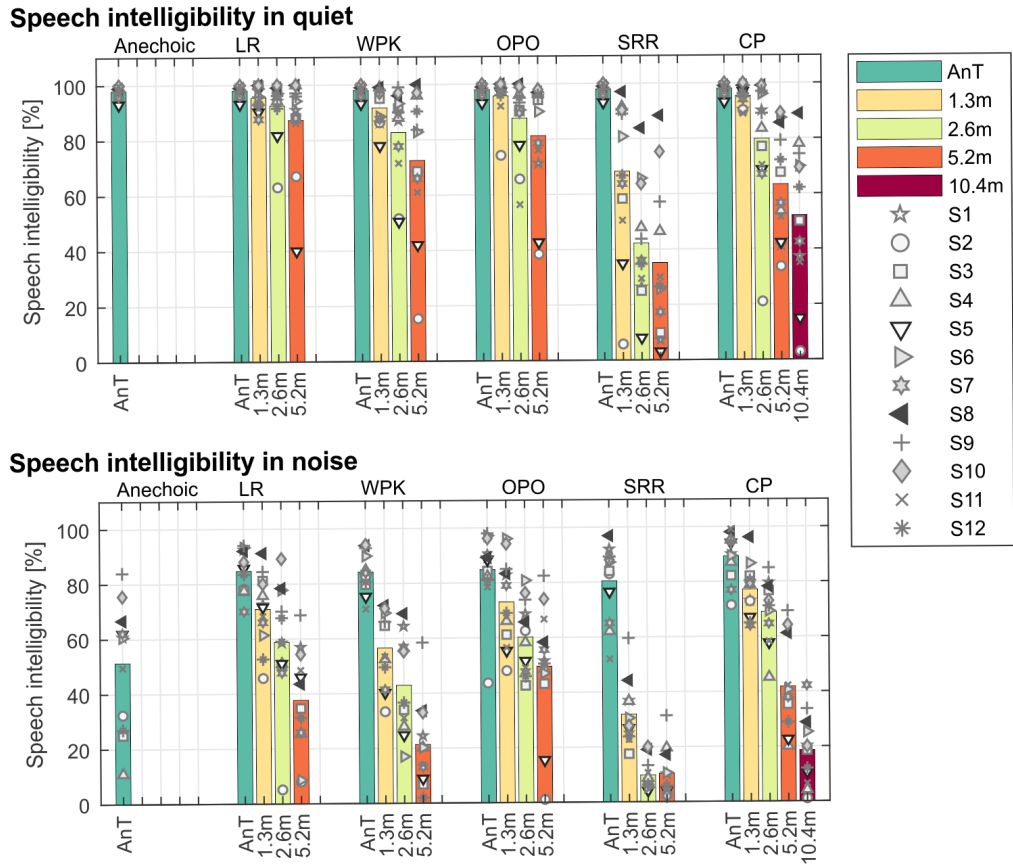


FIGURE 2.2: Speech Intelligibility scores obtained in quiet (top) and in noise (bottom). Each color represents a target-to-listener distance and each group of bars represents a room: Anechoic room, LR (lecture room), WPK (workplace kitchen), OPO (open-plan office), SRR (small reflective room) and CP (car park). Each bar represents the mean SI across participants. The noise level for all noisy conditions was 3 dB lower than the level producing 50% SI for an anechoic target (denoted as AnT) in the workplace kitchen noise.

with the impact of floor and ceiling effects, highlights that the SI in quiet relative to SI in noise is not simply a constant performance shift, nor does it change consistently as a function of distance across rooms.

A linear mixed-effects model fitted to the linearized SI scores (Sec. 2.2.5) revealed a significant effect of room [$F(4,315) = 24.38$; $p < 0.001$], a significant effect of distance [$F(1,315.02) = 213.88$; $p < 0.001$], a significant effect of noise [$F(1,315.02) = 183.26$; $p < 0.001$] and a significant interaction between noise and room [$F(4,315.02) = 2.48$; $p < 0.05$]. The interaction between distance, room and noise was also significant [$F(4,315.02) = 3.51$; $p < 0.01$]. The interaction between distance and room was not significant [$F(4,315) = 2.39$; $p = 0.05$], neither was the interaction between distance and noise [$F(4,315) = 0.05$; $p = 0.83$].

To understand if the different levels of reverberation inherent in the different noises (due to the different rooms) had an effect on SI, the same statistical model was used but only considering the noisy conditions with anechoic target speech (i.e.

the scores corresponding to the conditions depicted with a green bar at the bottom of Fig. 2.2). In this case, the model only included the room factor, which directly described the effect of the noise. The effect of noise was significant [$F(5,53.08) = 24.11$; $p < 0.001$]. Conducting pairwise comparisons with Turkey-corrections between the different noises with a series of t-tests (see Sec. 2.2.5) revealed that only the anechoic noise was significantly different from all the others ($p < 0.01$). None of the remaining comparisons revealed any significant differences.

To further understand the relative effects of reverberation and noise on individual SI performance, Fig. 2.3 shows the subjects' SI scores averaged across all quiet conditions as a function of their SRT. Note here that only the SRT presented an unbiased individual measure of the effect of noise, as all the other noisy conditions were tested at an SNR that depended on the individual SRT (i.e., the SNR was always 3dB below the individual SRT). Overall, the figure shows that participants who show high SI scores in quiet reverberant conditions show low SRTs [$R^2 = 0.73$], suggesting that listeners who tolerate noise well show also a better hearing in reverberation.

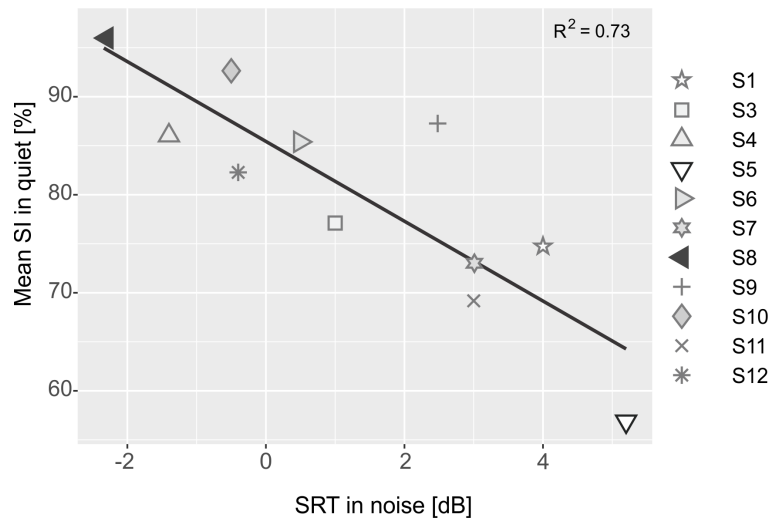


FIGURE 2.3: Speech intelligibility scores averaged across all quiet conditions as a function of the SRT obtained with anechoic target speech. Data of subject 2 has not been included, as not all the quiet conditions could be tested for this subject.

2.3.2 Questionnaire ratings

The questionnaire assessed the effects of room type, talker-to-listener distance, and noise on the subjective attributes of “listening effort”, “amount of reverberation”, and “noise distraction” for each of the 39 conditions tested in the SI experiment. Due to the large amount of data that resulted from this questionnaire, the actual data is not presented here, except for a particularly interesting sample shown in figure 2.4. Instead, the main trends in the data are reported based on a statistical analysis using a linear mixed-effects model (Sec. 2.2.5). Depending on the subjective attribute,

the model included the factors “room”, “distance”, and “noise”, with subjects as a random intercept.

The rating of effort generally increased with increasing talker-to-listener distance as well as in noise versus quiet, and varied across rooms in a way that was only partly correlated with their RT. All the main effects of distance [$F(1,322) = 37.72$; $p < 0.001$], room [$F(4,322) = 11.28$; $p < 0.001$] and noise [$F(1,322) = 94.15$; $p < 0.001$] were significant, but none of their interactions. Among all the potential interactions, the lowest p-value was observed for the interaction between room and noise [$F(4,322) = 1.98$; $p = 0.1$].

The attribute of distraction made only sense in noise, and the factor noise (i.e., quiet versus noise) was therefore not evaluated. The rating of distraction followed the same trends as those of effort, and both main effects of distance [$F(1,156) = 13.95$; $p < 0.001$] and room [$F(4,156) = 4.3$; $p < 0.01$] were significant, but not their interaction.

The rating of reverberation also increased significantly with talker-to-listener distance and varied across rooms. The main effects of distance [$F(1,335) = 88.03$; $p < 0.001$] and room [$F(4,335) = 56.70$; $p < 0.001$] were significant, but no significant effect of noise was observed [$F(1,335) = 1.5$; $p = 0.2$]. The interaction between distance and room was not significant.

To better understand the effect of the reverberation inherent in the different noises on ratings, as similarly done in Sec. 2.3.1 for the SI scores, Fig. 2.4 shows the effort and distraction ratings corresponding to the different noisy conditions with anechoic target speech. The plot shows that, on average, anechoic noise is more demanding in terms of effort, as well as more distracting, than reverberant noise in any room. Pairwise comparison (with Tukey correction) showed that effort ratings for anechoic noise were significantly different from all rooms ($p < 0.01$) except for the small reflective room [$t(50) = 2.71$; $p = 0.09$]. Regarding all the other paired comparisons, significant differences were observed only between the open-plan office and the small reflective room [$t(50) = -3.09$; $p < 0.05$]. For the rating of distraction, only the anechoic noise was significantly different from the other noises ($p < 0.01$) with the difference from the lecture room being just significant [$t(50) = 3.43$; $p = 0.05$]. No significant differences were observed between the distraction ratings of any of the other cases.

2.4 Discussion

In the following sections, the results from the SI test and questionnaire are both discussed separately for the quiet and noisy condition.

2.4.1 Results in quiet

In most of the rooms that were tested in this study, SI in quiet was not markedly compromised by reverberation at conversational distances (i.e., 1.3m). The only exception was the small reflective room, for which already at the closest distance of 1.3m the mean SI scores were below 80% and reached floor at a distance of 5.2m.

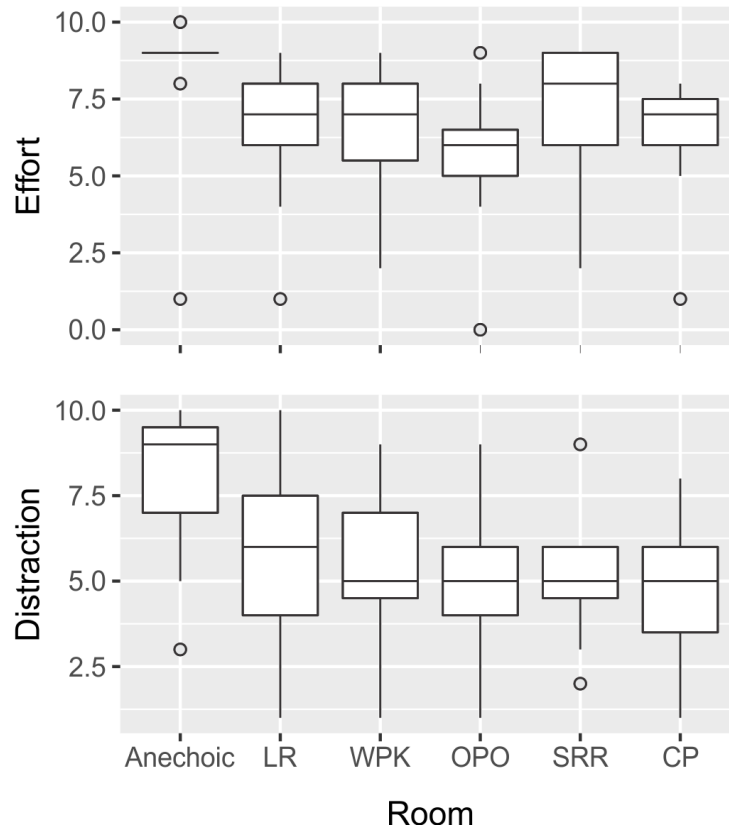


FIGURE 2.4: Median, 25th and 75th percentiles (hinges) of effort (top) and distraction (bottom) ratings obtained in the noisy conditions with anechoic target speech in the anechoic room (Anechoic), the lecture room (LR), workplace kitchen (WPK), open-plan office (OPO), small reflective room (SRR) and car park (CP).

Interestingly, the SI in this room was significantly lower than in the car park, even though the RT in the small reflective room ($RT = 1.55s$) was much shorter than in the car park ($RT = 2.42s$). A similar discrepancy between RT and SI was also observed between the workplace kitchen ($RT = 0.68s$) and the open-plan office ($RT = 0.96s$), where the workplace kitchen had the shorter RT but showed lower SI scores.

The above observations contradict the findings presented in previous studies with CI users in two different ways. First, reverberation is not as detrimental as suggested by previous work (Desmond, Collins, and Throckmorton, 2014; Hazrati and Loizou, 2012; Hu and Kokkinakis, 2014; Kokkinakis and Loizou, 2011; Kokkinakis, Hazrati, and Loizou, 2011; Poissant, Whitmal, and Freyman, 2006; Whitmal and Poissant, 2009). Second, SI does not necessarily decay monotonically with increasing RT as previously suggested (Desmond, Collins, and Throckmorton, 2014; Hazrati and Loizou, 2012; Hu and Kokkinakis, 2014; Kokkinakis and Loizou, 2011; Kokkinakis, Hazrati, and Loizou, 2011).

In order to understand the differences observed in the present and previous studies, it should be considered that the RT of a room depends on its volume and the absorption material (Sabine, 1993). Hence, the same RT can be obtained in a large

room with some absorption or in a smaller room with less absorption. Even though the two rooms have the same RT, the smaller room presents a higher reverberant pressure (page 88 in Jacobsen et al., 2013). This is most easily illustrated with the DRR, which enables direct comparisons between the reverberant pressure of different rooms at fixed talker-to-listener distances. The DRR can be approximated as:

$$DRR = 10 \log \left(\frac{\gamma 0.16V}{16\pi r^2 RT} \right), \quad (2.2)$$

with V the volume of the room in cubic meters, RT the reverberation time in seconds, γ the directivity index of the target source, and r the talker-to-listener distance in meters. Equation 2.2 was obtained from the quotient between the sound pressure squared of the direct sound and that of the stationary sound (page 88 in Jacobsen et al., 2013) and by using Sabine's equation to express the acoustic absorption as a function of the room volume and the RT.

In the case that a higher RT is obtained in a room with a larger volume, Eq. 2.2 predicts that the DRR at a given distance increases only if the proportional increase of volume is higher than the proportional increase of RT. In contrast, if the proportional increase of volume is lower than the proportional increase of RT, then Eq. 2.2 predicts a decreasing DRR. Hence, both decreasing and increasing DRRs can be found for increasing RTs obtained in larger rooms. With respect to Table 2.2, an example of lower DRRs found in a larger room can be seen by comparing the lecture room with the open-plan office. An example of higher DRRs found in a larger room can be seen by comparing the workplace kitchen with the open-plan office. According to the rooms employed here (Table 2.2), in case that a higher RT is obtained in a room with a larger volume, it is more common to observe increasing DRRs. In fact, the car park, which has the largest RT as well as the largest volume, has the highest DRR at all tested talker-to-listener distances.

In the case that a higher RT is obtained by reducing the overall absorption in the room while keeping the volume constant, Eq. 2.2 suggests that the DRR decreases with increasing RTs. With respect to Table 2.2, this refers to the lecture room, the workplace kitchen, and the small reflective room, which all have roughly the same volume but increasing RT, which results in a DRR that decreases for all talker-to-listener distances. At a distance of 5.2m, for example, the increase in RT results in a decrease in DRR of 1.9 dB from the lecture room ($RT = 0.46s$) to the workplace kitchen ($RT = 0.68s$), and in a further decrease of the DRR by 3.3 dB from the workplace kitchen to the small reflective room ($RT = 1.55s$).

As can be seen from the SI scores shown in Fig. 2.2, the negative effect of the RT on SI in quiet is higher in the second case than in the first case, indicating that an increase in RT that is accompanied by an increase in room volume is less detrimental to SI than an increase in RT that is achieved by reducing the amount of absorption inside a room. For example, at a talker-to-listener distance of 5.2m (Fig 2.2 upper panel), the drop in average SI scores from the workplace kitchen to the small reflective room

is much larger than from the open-plan office to the car park, even though the RT increases by roughly the same amount (i.e., by 0.87s and 0.96s, respectively).

Hence, the concept of reverberant pressure (or DRR) does not only explain why the RT is not a good predictor of SI, but also why, for a given distance, small reverberant rooms are more detrimental to SI than large reverberant rooms. Accordingly, the differences observed between the present and previous studies are most likely explained by the fact that most of the existing studies used a single (simulated or real) room with a rather small volume in which different RT values were obtained by altering the acoustic absorption of the room (Hazrati and Loizou, 2012; Kokkinakis, Hazrati, and Loizou, 2011; Kokkinakis and Loizou, 2011; Hu and Kokkinakis, 2014). In all these cases, SI will have decreased monotonically with the RT because the volume of the room was kept constant. As shown in the present study, this monotonic relationship between the RT and the achieved SI cannot be generalised to real rooms in which the volume often increases with increasing RT. Similarly, the rather strong detrimental effect of reverberation on SI that was observed in previous studies may be explained by the rather small reverberant rooms that were considered (Kokkinakis, Hazrati, and Loizou, 2011; Kokkinakis and Loizou, 2011; Hu and Kokkinakis, 2014; Whitmal and Poissant, 2009; Poissant, Whitmal, and Freyman, 2006) or by the large talker-to-listener distance (Hazrati and Loizou, 2012; Desmond, Collins, and Throckmorton, 2014) which will have both provided an exceptionally low DRR. For example, Kokkinakis, Hazrati, and Loizou (2011) considered a room with a volume of 76.8m³ and an RT of 1s, and observed mean SI scores as low as 20% at a rather close talker-to-listener distance of 1m. With respect to the present study, this case is more or less represented by the small reverberant room, which also showed by far the lowest SI scores of all the tested rooms. But even though this room was rather small (134m³) and with a high RT of 1.55s, average SI scores at 1.3m were still at 80% and thus, significantly higher than in Kokkinakis, Hazrati, and Loizou (2011) and many of the other related studies.

The overly strong impact of small, reverberant rooms on SI in CI users has already been highlighted by a number of studies (e.g., Galster, 2007; Gelfand and Silman, 1979; Nábélek and Robinette, 1978). However, none of the studies related this observation to the high reverberant pressure (or low DRR) that is found in these rooms, but rather to the high reflection density. Kressner, Westermann, and Buchholz (2018) provided the only study that, similarly to the present study, found that SI with CIs in realistic rooms is mainly compromised at far talker-to-listener distances. Moreover, by using the concept of the critical distance (i.e., the talker-to-listener distance at which the DRR is equal to 0 dB), they highlighted the risks associated with using a small single room with varying amounts of reverberation in laboratory-based tests, and argued that the room volume should be considered in addition to the RT.

It seems reasonable to assume that in most rooms in which humans communicate in their daily life, at least some form of absorption is applied (e.g., by carpets, suspended ceilings, cushioned furniture, or people) and that higher RT values are

accompanied by larger volumes. Following this assumption, the small reverberant rooms (either real or simulated) that were used in many of the existing studies had exceptionally low levels of acoustic absorption and hence, did not reflect the acoustic properties of the rooms commonly encountered in the real world. Hence, the corresponding SI results cannot be generalised and rather refer to a specific acoustic condition.

As also indicated by Eq. 2.2, the source directivity used for presenting the target speech material within the listening tests affects the DRR and thus, will have contributed to the differences between the SI scores reported in the literature and those obtained in the present study. This mismatch is likely larger in studies based on room acoustic simulations, as it is common practice to use omnidirectional sound sources (Whitmal and Poissant, 2009; Poissant, Whitmal, and Freyman, 2006). In the present study, the loudspeaker used during the RIR measurements (Sec. 2.2.2.2) did not present an omnidirectional pattern and thus, the DRRs obtained in the present study were relatively higher than those obtained with omnidirectional sources. However, even though the directivity of the applied loudspeaker was closer to the directivity of a human talker than an omnidirectional source, it was still an approximation and, as such, it did not perfectly match the directivity of a human talker. Hence, a slightly different detrimental effect of reverberation on SI would have been observed if the directivity of a real human talker had been considered.

Although the concept of the reverberant pressure (or DRR) helps to explain why the RT is not a good indicator of SI and why small reverberant rooms are more detrimental than any other type of room, the DRR alone is not a good indicator of SI either. As shown in Table 2.3, mean SI scores in quiet appear to be better correlated with the ratio of early to late reverberant energy, the C50 (Bradley, 1986), than with the DRR. This is consistent with the observations made in Kressner, Westermann, and Buchholz (2018), who found that the DRR does not directly correspond to SI scores with CIs and suggested that other measures such as the Speech Transmission Index (Houtgast, Steeneken, and Plomp, 1980) or the C50 may be more accurate predictors. As shown in Table 2.3, the same observation applies to the ratings of reverberation and effort. Interestingly, all the rooms in the present study that exhibit increasing DRR with increasing RT, also have decreasing C50 with increasing RT. This indicates that, although the reverberant pressure decreases for increasing RTs, the increase in RT results in lower C50s. Future work will need to conduct a more exhaustive analysis of the suitability of the C50 as an accurate predictor of SI, which would imply that CI recipients could benefit from early reflections in adverse situations. Likewise, given the high correlation between the participants' SRTs and their mean scores in quiet (Sec. 2.3.1), the use of the C50 could potentially be extended to the U50 (Bradley and Bistafa, 2002), an SNR-based metric that extends the concept of the C50 to allow the description of SI data obtained in quiet and noise under the same framework. Thereby, these metrics should take into account the signals that arrive at the subjects' speech processor worn during the listening test, which should include the directional

characteristics of the microphone.

TABLE 2.3: Coefficient of determination between mean scores in quiet, mean ratings of reverberation and effort in quiet and three different acoustic metrics: RT, DRR and C50

	RT (s)	DRR (dB)	C50 (dB)
Mean SI scores in quiet	$R^2 = 0.2$	$R^2 = 0.59$	$R^2 = 0.87$
Mean reverberation ratings in quiet	$R^2 = 0.36$	$R^2 = 0.54$	$R^2 = 0.92$
Mean effort ratings in quiet	$R^2 = 0.31$	$R^2 = 0.54$	$R^2 = 0.91$

2.4.2 Results in noise

As can be observed in the SI scores for the anechoic target speech shown in Fig. 2.2 (green bars) as well as in the effort and distraction ratings shown in Fig. 2.4, the most detrimental noise is the anechoic one, which is also the most modulated one, as its envelope is not smoothed by any reverberation (see Sec. 2.1). This is in general agreement with other studies, which have found that SI with CIs is worse with fluctuating noises than with steady-state noises (Fu and Nogaki, 2005; Nelson et al., 2003; Qin and Oxenham, 2003). However, for the anechoic talker no major differences were observed between all the reverberant noise conditions. This may be in part due to the fact that the SI scores in all these conditions were already close to ceiling and thus, not in a very sensitive point of the psychometric function.

In the case that the broadband level of the noise in a given room would fully describe the impact of the noise on SI, the rooms sorted by SI in quiet would coincide with the rooms sorted by SI in noise, as all noises were adjusted to the same subject-specific broadband level (see Sec. 2.2.2.2). As can be seen in Fig. 2.2, this is not necessarily the case. For example, while the lecture room appears to be the most favourable room in quiet, the most favourable room in noise is the open-plan office. Hence, there is either an interaction between the reverberation of the target speech and the noise, or factors other than the broadband level of the noise may have affected SI. This may include the modulation of the noise but also its frequency spectrum.

However, a comparison of the effect of the reverberation of the different rooms on the noise from the results plotted in Fig. 2.2 is not straightforward, because the effect of noise on SI for any room and talker-to-listener distance depends on where in the psychometric function its quiet counterpart is located. For example, while Fig. 2.2 may suggest that the reduction in SI due to noise is greater for the workplace kitchen than in the lecture room, it is not possible to say whether that would also be

true if the scores in quiet in the two rooms had been matched. While this question remains unanswered here, several broad observations can be made by comparing reverberant conditions in which, either in quiet or in noise, the scores are comparable. For example, the workplace kitchen and the car park lead to similar SI scores in quiet at 2.6m. The fact that scores in noise in the workplace kitchen are much lower than those achieved in the car park indicates that the noise in the workplace kitchen is more detrimental to SI than the noise in the car park. Another example can be seen in the lecture room and in the car park where SI scores at 5.2m in noise are comparable. The fact that in quiet, scores obtained in the lecture room are much higher indicates that the noise in the lecture room is more detrimental than the noise in the car park.

The fact that the room with the highest SI scores in noise does not always correspond to the room with the highest scores in quiet raises the question of what is the room that leads to the best trade-off between quiet and noise. Figure 2.5 shows the mean SI scores obtained in noise as a function of the SI scores obtained in quiet (left panel), as well as their corresponding mean effort ratings reported in the questionnaire (right panel). Each condition is indicated by a number, which corresponds to the ID column presented in Table 2.2. The most favourable reverberation conditions are those that present high scores in both quiet and noise, which can be found by identifying the closest points to the upper-right corner. As the distance from the upper-right corner increases, less favourable reverberant conditions are found, with the least favourable conditions being closest to the bottom-left corner. Thus, when considering the quiet and noise conditions at the same time, the car park at 1.3m (number 13), the open-plan office at 1.3m (number 7) and the lecture room at 1.3m (number 1) are the most favourable reverberation conditions. Interestingly, while SI in quiet for these conditions is almost as good as in anechoic conditions (number 0), in noise it is much higher. Most favourable conditions in terms of SI scores are in agreement with the mean effort ratings reported in the questionnaire (Fig. 2.5, right panel).

Hence, when considering quiet and noisy conditions, the most favourable listening conditions for CI users appear to consist of short distances in rooms with large volumes (i.e., the car park and the open-plan office) or in smaller rooms with some reverberation (e.g., the lecture room). Under those conditions, SI in quiet is not compromised by target reverberation due to the high DRR. And in noise, reverberation reduces the modulation of the noise, thus making it less detrimental than anechoic noise. The combination of these two factors leads to better and less effortful speech intelligibility. Nevertheless, it is important to note the relevant limitations of the present study. First, the conclusions drawn here apply to modulated noise. Clearly, there are several factors that have an impact on the modulation depth of the noise signals, such as the distance to the listener or the number of noise interferers. Second, target and masker levels were normalised. Reverberation has an effect on these levels and therefore, different conclusions may be obtained when comparing reverberant conditions at their non-normalised levels. Third, most of the advanced

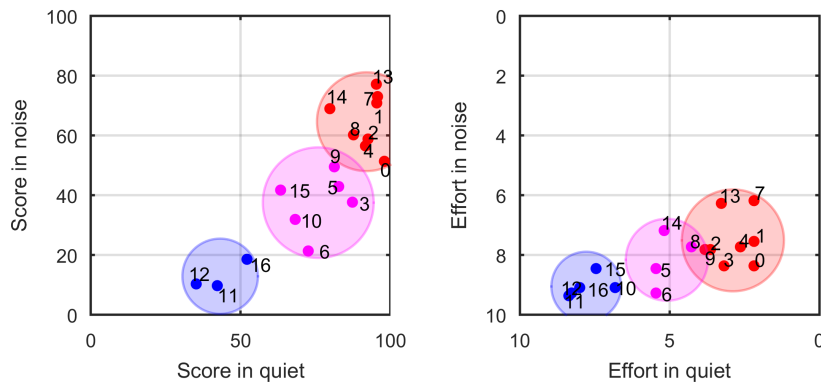


FIGURE 2.5: Mean SI score in noise as a function of mean SI score in quiet (left) and mean effort rating in noise as a function of mean effort rating in quite (right). The numbers represent a reverberant condition, as specified in Table 2. The number 0 represents the anechoic condition (no room). For illustrative purposes, the data have been clustered into three groups by means of the k-means algorithm. Clustering process in right and left are independent to one another.

features of the applied CIs were switched off during testing. Hence, it is unclear if the same observations apply to the case when these feature are turned on, and most likely interact with the different reverberant stimuli.

2.5 Conclusions

The present study evaluated speech intelligibility in quiet and noise under a number of realistic reverberant conditions that CI recipients may experience in their daily lives. Results in quiet show that in most cases speech intelligibility is not compromised by reverberation at conversational distances of less than a few meters. Nevertheless, small reflective rooms have been shown to affect SI more than any other type of room. The results and conclusions presented here are in contrast to most of the prior literature in two ways. First, reverberation in realistic scenarios affects speech intelligibility for CI users less than suggested previously, and second, speech intelligibility does not necessarily decay monotonically with increasing reverberation times. These contradictions can be explained by the fact that prior studies used rather small and highly reverberant rooms (e.g. 76.8m^3), and therefore evaluated rooms that are more detrimental to speech intelligibility than commonly encountered in the real world. Moreover, these studies realised different reverberation conditions by altering the absorption material while keeping the volume of the room constant, which lead to SI scores that could be uniquely explained by the RT.

Results in noise indicate that the reverberation of modulated noise (here 4-talker babble) may actually contribute positively to speech understanding, and that anechoic noise is the most detrimental to SI. This result can be attributed to the fact that reverberation reduces the noise modulation. Results in quiet and noise evaluated together suggest that short talker-to-listener distances combined with large volumes

or smaller volumes with some reverberation yield the best outcomes in terms of SI and self-reported listening effort.

Chapter 3

Validation of existing room acoustic criteria for predicting speech intelligibility with cochlear implants

Abstract

Most existing models for predicting speech intelligibility in cochlear implant (CI) recipients are purely signal-driven (i.e. based on the speech signal) and do not allow any conclusions on the effect of the basic room acoustic parameters on speech intelligibility performance in reverberant environments. Although room acoustic measures (i.e. based on the room impulse response) such as the U50 (an extension of the clarity measure C50) or the speech transmission index (STI) are commonly used to predict speech intelligibility in rooms for normal-hearing listeners, their relevance for CI recipients is still unclear. This study investigates the suitability of the U50 to describe speech intelligibility data of twelve cochlear implant users tested unilaterally over a wide range of realistic rooms and talker-to-listener distances in both quiet and noise. Results in quiet show that speech intelligibility can be accurately predicted by the U50. The resulting U50-based room acoustic criteria of speech intelligibility for CIs, as well as acoustic design limitations and potential alternatives, are discussed alongside existing criteria, usually oriented to people with normal hearing. Additionally, in order to understand the conflicting results obtained across studies, room parameters most critical to speech intelligibility in quiet are discussed. Results in noise show that the temporal smearing effect of reverberation flattens the envelope of the noise signal thereby reducing its impact on speech intelligibility. As U50-based predictions do not account for this effect, speech intelligibility predictions in noisy conditions warrant further improvements. Predictions based on the speech transmission index lead to the same conclusions as with the U50 in both quiet and noise. Based on the observed dependence of speech intelligibility on the modulation of both target speech and noise, models for CI recipients can potentially benefit from predictions obtained in the modulation domain.

3.1 Introduction

The negative impact of reverberation on speech intelligibility (SI) is greater for people with cochlear implants (CI) than for normal hearing (NH) listeners (Kokkinakis, Hazrati, and Loizou, 2011; Kressner, Westermann, and Buchholz, 2018). However, there is still debate regarding the reverberant conditions in which CI recipients have difficulties to understand speech even in quiet. A number of studies have suggested that SI in quiet can already be challenging in apparently moderate reverberant conditions such as with reverberation times (RT) of 0.6 seconds and talker-to-listener distances as close as one meter (e.g. Kokkinakis, Hazrati, and Loizou, 2011). Other studies have not been able to reproduce such strong impact of reverberation suggesting the effect of reverberation may not be as detrimental as previously suggested (e.g. Kressner, Westermann, and Buchholz, 2018). In Chapter 2, SI was found to be greatly affected by small reverberant rooms at typical conversational distances as well as at long talker-to-listener distances in large reverberant rooms. Moreover, studies have demonstrated that the temporal smearing effect of reverberation on the envelope modulations of the noise signal may have an impact on speech intelligibility, a finding that is in line with the previously reported lower SI performance of CI users with modulated noises (Fu and Nogaki, 2005; Qin and Oxenham, 2003; Nelson et al., 2003). The disparity of SI results obtained across different studies raises the question of which room parameters are the best predictors of the impact of reverberation on CI performance.

In contrast to CI users, the main room acoustic parameters affecting SI with NH listeners are well understood. Room acoustic criteria for speech intelligibility are usually based on the Speech Transmission Index (STI, Houtgast, Steeneken, and Plomp, 1980; IEC, 2003) or on the clarity-50 (C50, ISO 3382-1, 2009), usually referred to as U50 when the effect of noise is included (Nijs and Rychtáriková, 2011; Bradley and Bistafa, 2002; Bradley, 1986; Bistafa and Bradley, 2000). These metrics can be approximately expressed as a function of more basic room parameters such as room volume, RT or talker-to-listener distance, which provides a good understanding of the room parameters most relevant to SI (e.g. Houtgast, Steeneken, and Plomp, 1980; Nijs and Rychtáriková, 2011; Bradley and Bistafa, 2002; Bradley, 1986; Bistafa and Bradley, 2000). Because these metrics are standardized measures that are already part of the acoustic design in most acoustically-treated venues, and given the valuable insights they provide into the critical room parameters affecting SI, the benefit of validating these metrics for CI users is twofold: (1) to evaluate whether the acoustics of a given room are amenable to people who rely on CIs, and (2) to assist in the design of signal processing algorithms aimed to improve speech intelligibility in (noisy) rooms, including beamformers, scene/room classifiers and de-reverberation processing.

This study therefore aims to (1) provide a first step towards a room acoustic SI prediction model that is based on the concept of the U50 and (2) to use that model to understand the critical room acoustic parameters and their effect on SI performance

in CI users. This will be carried out by using the data from Chapter 2. The U50 is used here instead of the STI, as it provides some insight into the ability of the hearing system to integrate early reflections, a mechanism that has not yet been well documented in CI users. In addition, the U50 is a rather easy to interpret metric that combines a classic signal-to-noise ratio (SNR) measure, accounting for the effect of the noise, along with the C50, which considers the effect of reverberation. This, together with the fact that the U50 values are not constrained (i.e., they have no maximum or minimum values), as opposed to the STI, makes it easier to fit performance intensity functions. Aside from the fundamental differences between these two metrics, the high correlation between them has led researchers to conclude that they predict SI with comparable accuracy (Bradley, 1986). Hence, the results found here by using the U50 metrics may be directly applicable to the STI.

The U50 is a measure that has been used extensively as a room acoustic descriptor of SI in NH listeners (e.g. Nijs and Rychtáriková, 2011; Bradley and Bistafa, 2002; Bradley, 1986; Bistafa and Bradley, 2000). Generally speaking, speech intelligibility is regarded as "good" for U50 values ranging between 1.5 dB and 6.5 dB. Above this value, SI is regarded as "excellent" (Nijs and Rychtáriková, 2011). However, it is unclear how far these general guidelines apply to CI users. One of the motivations of the present study is therefore to verify the performance of the U50 as a room acoustic descriptor so that its value can be related to the same quality attributes but for CI recipients. This will hopefully put the difficulties faced by CI users into perspective, as the use of the U50 will enable a comparison between CI and NH listeners under the same framework.

Despite the progress of SI modeling strategies on the ability to predict the effect of noise and reverberation on SI with CIs, most current SI predictors are signal-driven models that, as opposed to models based on room acoustics, do not aim at providing a good understanding of basic parameters affecting SI (e.g. talker-to-listener distance, room volume, RT) as they rather focus on the minimization of the prediction error. In Goldsworthy and Greenberg (2004), four different speech-based STI methods were presented as good predictors of SI of nonlinearly processed speech. Based on the similarities between the procedure involved in the STI calculation and in the processing of speech for CIs, the authors concluded that these predictors could potentially be good predictors of SI with CIs (Goldsworthy and Greenberg, 2004). One of the measures presented in Goldsworthy and Greenberg (2004), the *envelope regression method*, was later used in SI tests conducted with NH and vocoded speech and showed highly accurate predictions (Poissant, Whitmal, and Freyman, 2006; Whitmal and Poissant, 2009). Another method, called *short-term objective intelligibility* (STOI), was proposed in Taal et al. (2011), and is based on the short-term correlation coefficient between clean and degraded speech. Motivated by the need of finding non-intrusive methods, as they can be applied in situations where the clean signal is unknown, Chen, Hazrati, and Loizou (2013) presented a predictor called *ModA*, which is based on the modulation spectrum and devised for SI prediction of CIs

in reverberation. Comparisons conducted in their study concluded that *ModA* outperforms the *normalized covariance measure*, one of the four methods previously presented in Goldsworthy and Greenberg (2004). Another non-intrusive measure was proposed in Santos et al. (2013) as a CI-oriented version of an existing method (Falk, Zheng, and Chan, 2010), which they named *speech-to-reverberation modulation energy ratio* (with the variation being called *SRMR - CI*). Comparisons between *ModA* and an updated version of the *SRMR - CI* showed that the latter predicts SI more accurately under a number of noisy, reverberant and processed conditions (Santos and Falk, 2014). A following study comparing twelve SI predictors concluded that *STOI* and *SRMR - CI* lead to the best predictions of CI intelligibility of reverberant, noisy and processed conditions (Falk et al., 2015). Yet another method, the output SNR (OSNR) was presented in Watkins, Swanson, and Suaning (2018), a method capable of accounting for the dependence of SI on the presentation levels. The fundamental difference of the present study with respect to the existing SI models is that SI modeling is here motivated by the attainment of a better understanding of the main room parameters that affect SI in CI users, which is not provided by the existing signal-driven predictors.

Aside from studies devoted to SI modeling, room acoustic parameters affecting SI with CIs in quiet have been most commonly described in terms of the reverberation time (RT, e.g. Kokkinakis, Hazrati, and Loizou, 2011; Desmond, Collins, and Throckmorton, 2014). Only few studies have focused on more complete metrics that can be generalized to other rooms and talker-to-listener distances. For example, Poissant, Whitmal, and Freyman (2006) and Whitmal and Poissant (2009) demonstrated that the speech-based STI could successfully predict their average SI data. However, these results were obtained with vocoded speech signals presented to NH listeners, and the extent to which the results can be generalised to CI users remains unclear. Similarly, in a study conducted with CI recipients under a range of realistic rooms and source-to-listener distances, Kressner, Westermann, and Buchholz (2018) drew similar conclusions as in Poissant, Whitmal, and Freyman (2006) and Whitmal and Poissant (2009), suggesting that both the STI and the C50 could potentially be good candidates for SI predictions in CI recipients across various rooms. However, a quantitative assessment of the suitability of the STI and the C50 was not conducted in this study, likely due to the fact that most SI scores were near or at ceiling (Kressner, Westermann, and Buchholz, 2018). Another study that found a strong correlation between the C50 and mean SI scores in reverberant quiet conditions is described in Chapter 2. However, only a simple linear regression analysis was conducted between the C50 and average SI values. It therefore remains unclear whether the C50 measure can be used at an individual level, and whether the addition of the noise component has an effect on the prediction accuracy.

Neither the STI nor the U50 consider the effect of noise modulation on SI. However, previous studies evaluating SI with CIs have shown that, for a given SNR, modulated noises have a more detrimental effect on SI than non-modulated noises (Fu

and Nogaki, 2005; Qin and Oxenham, 2003; Nelson et al., 2003). Studies have suggested that this is likely due to the limited fine-structure cues available to CI recipients, which results in the inability to identify changes in the temporal fine structure in the dips of a noise signal (Hopkins and Moore, 2009) and thereby hinder auditory stream segregation. In Chapter 2, the effect on SI of four pairs of interfering talkers surrounding the listener was shown to be stronger in anechoic conditions than in any of the five rooms included in the study (with low, moderate and high reverberant conditions). As discussed in the paper, these differences may be explained by the fact that reverberation makes the envelope of the noise signal more shallow (Houtgast, Steeneken, and Plomp, 1980), thereby reducing the negative impact of noise modulations on SI. In line with these findings, the present study focuses on the effect of noise reverberation on SI. This is in contrast to previous studies, which have mostly evaluated the effect of generic, anechoic noises (Hazrati and Loizou, 2012) or used vocoded speech with NH individuals (Poissant, Whitmal, and Freyman, 2006; Whitmal and Poissant, 2009).

Many of the previous studies either used room-acoustic simulations, CI speech simulations, or a combination of both (e.g. Kressner, Westermann, and Buchholz, 2018; Desmond, Collins, and Throckmorton, 2014; Whitmal and Poissant, 2009; Poissant, Whitmal, and Freyman, 2006). Studies that used actual recorded Room Impulse Responses (RIRs) when testing CI recipients mainly obtained different reverberation conditions by altering the absorption material of a single, small room (e.g. Kokkinakis, Hazrati, and Loizou, 2011). In an attempt to obtain more ecologically-valid results, and to conduct a more exhaustive investigation of the different reverberation conditions that a person may encounter in their daily life, the data considered in the present study was obtained by presenting CI recipients with three-dimensional representations of the sound field of five real rooms with distinct RTs at three different talker-listener distances (Chapter 2). Hence, the present study contributes to the field by conducting a systematic investigation of the suitability of the U50 as a SI criterion of rooms for the particular case of CIs.

3.2 Methods

As mentioned above, the goal of the present study is to evaluate the suitability of the U50 as a room acoustic descriptor of SI with CIs. This evaluation is conducted by fitting U50 values to SI scores expressed as percentage of morphemes correctly understood. Because the SI scores were presented in a previous study (Chapter 2), this section focuses mainly on the calculation and the fitting of the U50 values. The calculation of the U50 is based on simulations of the stimuli received by the speech processor of the subject's CI used during the listening tests, and the fitting is based on a non-linear mixed effects model using a sigmoidal model function.

3.2.1 Stimuli and speech intelligibility data

The suitability of the U50 to predict SI in CI recipients was evaluated here using the extensive data set measured in Chapter 2. In brief, speech intelligibility of twelve post-lingually deafened CI recipients was tested unilaterally in 39 different acoustic conditions. BKB-like sentences (Bench, Kowal, and Bamford, 1979) were convolved with multichannel room impulse responses (RIRs) and presented to the subjects via a spherical array of 41 loudspeakers. The RIRs were obtained in five rooms with distinct RTs at three or four talker-to-listener distances (see Table 3.1) using a 62-channel hard-sphere microphone array. The recorded RIRs were then decoded into loudspeaker signals using the Higher-order Ambisonics (HOA) method (see Weisser et al., 2019 for details). In each acoustic condition, SI was tested in both quiet and noise. The noise consisted of four two-talker dialogues recorded in anechoic conditions and convolved with RIRs obtained in each of the five rooms at positions surrounding the listener at either 2m (two dialogues) or 2.6m (remaining two dialogues) from the listener. The positions of the eight interfering talkers, relative to the listener's tested ear, were the same regardless of the room, the talker-to-listener distance and the subject. In no case was the listener located at the axis of incidence (i.e. in front) of the interfering noise sources. For reference purposes, SI was also measured for anechoic speech in quiet, in anechoic noise, and in each of the reverberant noises of the five rooms. This resulted in 39 different conditions in total.

During the listening tests, the overall level of the target speech including reverberation was fixed at 60 dB SPL. The overall level of the noise, also including reverberation, was kept fixed throughout the entire experiment but varied across subjects, a decision that was taken to minimise ceiling and floor effects. In order to determine the noise level used during the test, each subject's speech-reception-threshold (SRT) for 50% correct (Keidser et al., 2013) was measured two times at the beginning of the tests and their values averaged. The SRTs were measured with anechoic speech in the aforementioned noise layout located in the workplace kitchen (see Table 3.1). The noise level was then determined as 60 dB SPL - SRT₅₀ - 3dB. Presentation levels were all calibrated with an omnidirectional microphone located at the centre of the loudspeaker array.

During the tests, the following speech processor features were disabled: Microphone directivity, Automatic Dynamic Range Optimization (ADRO), SNR-NR, Spatial-NR, SCAN, WhisperTM and WNR. Out of the two microphones located in the speech processor, only the microphone located at the front was enabled, which led to an omnidirectional directivity response. Out of the pre-processing algorithms explicitly designed to improve SI in noise, only Automatic Sensitivity Control (ASC) was enabled. ASC progressively reduces the sensitivity of the microphone as the level of the noise increases (Wolfe et al., 2015) to ensure that the dynamic range of the input signal is not compromised by the compressor. In fact, if ASC was not enabled, signals reaching 65 dB SPL or higher would have been clipped (Gifford and Revit,

TABLE 3.1: Reverberation time (RT), room volume, critical distance, DRR and C50 for each of the rooms and distances included in this study. RT and DRR were obtained from the RIR as simulated with an omnidirectional microphone located at the centre of the loudspeaker array. The critical distance was derived from the decay over distance of the DRR. C50 values were obtained from the speech signals simulated at the output of the BTE microphone of the subject's speech processor. RT, DRR and C50 values were calculated as the mean across third octave bands from 125Hz to 8kHz.

Room	RT (s)	Room volume (m ³)	Critical distance (m)	Distance (m)	DRR (dB)	C50 (dB)
Lecture Room	0.46	164	2.78	1.3	6.9	14.4
				2.6	-0.6	10.6
				5.2	-4.6	8.2
Workplace kitchen	0.68	164	1.77	1.3	2.7	8.6
				2.6	-3.8	5.5
				5.2	-6.5	4.4
Open-plan office	0.96	446	2.49	1.3	5.2	12.6
				2.6	-1.5	9.6
				5.2	-4.5	7.9
Small reflective room	1.55	134	1.18	1.3	-0.3	2.5
				2.6	-6.2	0.6
				5.2	-9.8	-1.1
Indoor car park	2.42	> 5700	3.34	1.3	7.2	11.4
				2.6	0.4	6.1
				5.2	-4	2.9
				10.4	-6.1	1.8

2011). The speech processor used during the tests included all the subject's individual fitting parameters (stimulation rate, number of maxima selection, T and C levels) as provided by their clinical audiologist.

3.2.2 BTE signal simulation

To simulate both the (reverberant) target speech and noise signals that were picked up by the front microphone of the subject's speech processor for all the tested acoustic conditions, first, a HATS wearing the behind-the-ear (BTE) speech processor used during the experiments was located at the center of the loudspeaker array. A set of 41 impulse responses was obtained, each of them describing the acoustic path from a loudspeaker to the BTE microphone. For each reverberant condition, each of the 41 loudspeaker impulse responses was convolved with its corresponding channel of the 41-channel RIR obtained in the different rooms and decoded into the HOA format (see above). The 41 channels were superimposed (i.e., added up) to obtain a single-channel RIR. By convolving this single-channel RIR with anechoic speech, the target and interferer signals received by the BTE microphone of the subject's speech processor during the listening tests was simulated for each acoustic condition.

In order to calculate the U50, at this stage, the derived single-channel RIRs were truncated between 0 and 50ms to describe the direct sound plus early reflections (DSER) component and from 50ms onward to describe the late reverberation (LR) component of the RIR. The non-truncated RIRs were also simulated and used later for calibration purposes, as the presentation levels during the listening tests corresponded to the broadband levels of the target and noise signals including reverberation.

For the target speech, 1264 anechoic sentences available in the applied BKB-like speech corpus were concatenated and transformed into BTE microphone signals for each acoustic condition separately using the simulation method described above. For each masker, the speech of the eight talkers involved in the four dialogues were independently transformed into reverberant BTE microphone signals and added up thereafter. In the case of anechoic target speech or maskers, the BTE signals were obtained directly by convolving the anechoic signal with the impulse response measured from the given playback loudspeaker to the BTE microphone located on HATS. The BTE microphone signal simulations for the target signals in all 39 acoustic conditions were derived separately for the entire RIRs, the DSER components, and the LR components, as well as for all the individual masker levels at which the 12 CI recipients were tested.

In order to verify the BTE microphone signal simulations, true dB SPL presentation levels were measured with the BTE worn by the HATS (located in the anechoic chamber) and compared to the simulated values. The comparison was conducted for a subset of BKB-like sentences randomly selected and convolved with the non-truncated RIRs. By doing this, it was ensured that sound pressure levels used during the analysis were exactly the same as the ones presented in the actual experiment.

3.2.3 Calculation of the U50

Similarly to a SNR, the U50 is expressed as:

$$U50 = 10 \cdot \log\left(\frac{DSE R}{LR + N}\right), \quad (3.1)$$

where *DSE R* is the combined power of the direct sound and early reflections components of the RIR convolved with the target speech, *LR* is the power of the late reverberation component of the RIR convolved with the target speech, and *N* is the total power of the noise (see Chapter 4, page 104 in Kates, 2008). The power of the different RIR components were derived here from the reverberant target speech signals simulated at the BTE microphone of the subject's speech processor, as described in Sec. (3.2.2). Equation (3.1) was calculated in third octave bands and their values averaged. Thereby, only the frequency bands between 125Hz and 8kHz were considered to account for the frequency range most relevant to speech understanding. Note that the target and masker stimuli applied in the listening tests were all calibrated using their broadband levels measured with an omnidirectional microphone in the centre of the playback loudspeaker array (Sec. 3.2.1). Considering that the U50 was derived from the simulated BTE microphone signals of the subject's speech processor and averaged across frequency bands after logarithmic compression to dB values (see Eq. 3.1), the U50 varied significantly across conditions.

3.2.4 Temporal modulations of the noise

Given that the broadband noise level was normalized for each subject across all noise conditions, the main characteristics that changed as a function of reverberation (i.e., as a function of talker-to-listener distance and type of room) was its frequency spectrum as well as its modulation spectrum. The effect of the frequency spectrum on speech intelligibility is already addressed by the frequency dependency of the U50 (Sec. 3.2.3). In contrast, the potential effect of the reverberation on the amplitude modulations of the noise on speech intelligibility is not captured by the U50. To overcome this potential limitation of the U50, the modulation spectrum was derived for each noise considering the simulated BTE microphone signals received by the subject's speech processor. Each noise signal was first filtered by a complex, fourth-order Gammatone filterbank (see Hohmann, 2002) and the Hilbert envelope derived for each frequency channel by taking the absolute value of the filtered output signal. Each frequency channel was then analyzed by a modulation filterbank with octave-wide bands at center frequencies between 1 and 8Hz. The resulting modulation spectrum was then normalized separately in each auditory frequency channel to the overall power within that auditory channel. After this calculation was conducted in all rooms, all the modulation spectra were arbitrarily normalized to the maximum

value across all frequency bands, modulation bands and rooms. The final modulation spectra were averaged across all auditory frequency channels and are shown in Fig. (4.9) for the six different noises.

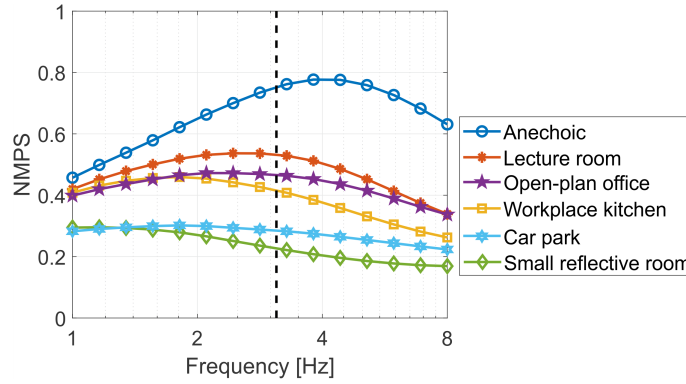


FIGURE 3.1: Normalized Modulation Power Spectrum (NMPS) of the noise signals in each of the different rooms. The modulation frequency that corresponds to the maximum of the NMPS for the anechoic target speech is indicated by the dashed line. See text for details.

The same modulation analysis was conducted to evaluate the most prominent modulation frequency of the anechoic target speech, which was 3.1 Hz and is shown as a dashed vertical line in Fig. 4.9. Motivated by the concept of the SNR in the modulation domain (Jørgensen and Dau, 2011), it is assumed in the following that the modulation depth of the noise at this specific modulation frequency is a proxy measure for how detrimental the noise modulations are for SI in CI recipients. Hence, it is assumed that highly modulated noises lead to low SNRs, making them more effective maskers. As can be seen in Fig. 4.9, the anechoic noise is the most modulated followed by the lecture room, workplace kitchen, open-plan office, car park and the small reflective room.

3.2.5 Statistical model

As described briefly in Sec. (3.2.1) and in detail in Chapter 2, SI scores were measured in 39 acoustic conditions, which resulted in a wide range of performance within and across subjects. It is assumed here that any systematic within-subject variation of the SI performance across acoustic conditions can be predicted successfully by the U50 (see Sec. 3.2.3). If this is the case, then the SI scores as a function of the U50 can be successfully described by a simple psychometric (or sigmoid) function and its parameters estimated from the measured data. To evaluate this assumption, a non-linear mixed effects model was applied here that utilises a sigmoidal model function with SI scores as the dependent (or output) variable and U50 as the independent (or predictor) variable.

The analysis of the SI data was divided into seven categories. The first category is referred to as *quiet* and included all the quiet conditions. This was necessary because the quiet data could not be analyzed in a per room basis due to ceiling effects (i.e.,

the SI scores for most subjects and conditions was at or near ceiling). The other six categories corresponded to the data in noise obtained in each of the six rooms (anechoic, lecture room, workplace kitchen, open-plan office, small reflective room and car park). Speech intelligibility of the i 'th subject in the j 'th category was represented as a function of the U50 by means of the following sigmoid function:

$$SI_{ij} = \frac{100}{1 + e^{\frac{-4k_i}{100}(U50 - \overline{SRT}_{ij})}} + \epsilon_{ij}. \quad (3.2)$$

where k_i represents the slope of the curve (in %/dB) at 50% SI and \overline{SRT}_{ij} is the U50 at 50% SI, which corresponds to the estimate of the SRT of the subject i in the category j . As this study focuses primarily on the effects of room acoustics on SI, the model included both a random slope k and an \overline{SRT} for each subject. Regarding fixed effects, the model included a separate \overline{SRT} for each category but a single (common) slope k . Further details about the statistical model can be found in Sec. (3.A).

The U50 values were fitted to the data in R (version 3.4.1) using the *nlme* package (version 3.1-131, Pinheiro and Bates, 2000). When reported, the approximate confidence intervals of the estimated parameters were obtained using approximate distributions for the maximum likelihood estimates and the restricted maximum likelihood estimates (Pinheiro and Bates, 2000). Prediction intervals, depicted as 95% percentiles of predictions, were obtained from resampling multiple (1000) times the fixed effects of the model. This technique assumed that the fixed effects were multivariate normal, and random-effect (individual level) parameters were ignored.

The fitting error between the measured and modeled data was evaluated by means of the root mean square (RMS) error, which is expressed as:

$$RMS\ error = \sqrt{\frac{1}{N-d} \sum_{k=1}^N (SI_{ik} - \overline{SI}_{ik})^2}, \quad (3.3)$$

where N indicates the number of conditions, d indicates the degrees of freedom (here $d = 2$), SI_{ik} is the speech intelligibility score obtained by subject i in the k 'th condition and \overline{SI}_{ik} is its predicted counterpart. The evaluation of the *RMS error* was evaluated at an individual level i and distinguished between quiet and noise conditions. The evaluation of noise conditions was not conducted in a room basis but rather, all noise conditions together (as opposed to the fitting procedure; see index j in Eq. 3.2). Hence, N equaled 17 for quiet conditions and 22 for noise conditions.

During the analysis, the measured SRTs (see Sec. 3.2.1) will be referred to as SRT_N . Note that this value corresponds to the U50 at the output of the BTE speech processor's microphone (see Sec. 3.2.2) at which the subject obtained 50% correct SI, and was calculated following the steps explained in Sec. (3.1). Estimated SRTs will be denoted with a bar as \overline{SRT} . Whether the estimates are at the individual or at the group level will be specified in each case.

3.3 Results

This section is divided into four subsections. First, the U50 is evaluated in quiet conditions. Second, the results of a room-dependent U50 fitting in noise are presented. Third, in order to test the assumptions of the relative effects of noise and reverberation underlying the U50, the subject-dependent \overline{SRT} in quiet (i.e. the SRT estimated from the fitted curves obtained in quiet conditions) is compared to the SRT of the subject (SRT_N), which was obtained with anechoic target speech in the workplace kitchen noise (see Table 3.1). Fourth, subject effects are investigated by evaluating the subject-specific U50 fitting parameters, i.e., by considering the random effects of the model described in Sec. 3.2.5.

3.3.1 Effect of reverberation in quiet

Figure 3.2 shows the SI scores obtained by each subject in the 17 quiet conditions (filled circles). The psychometric (sigmoid) functions fitted to the subjects' individual SI scores in quiet are shown by the solid lines, whereas the dashed curves show the prediction at the group level.

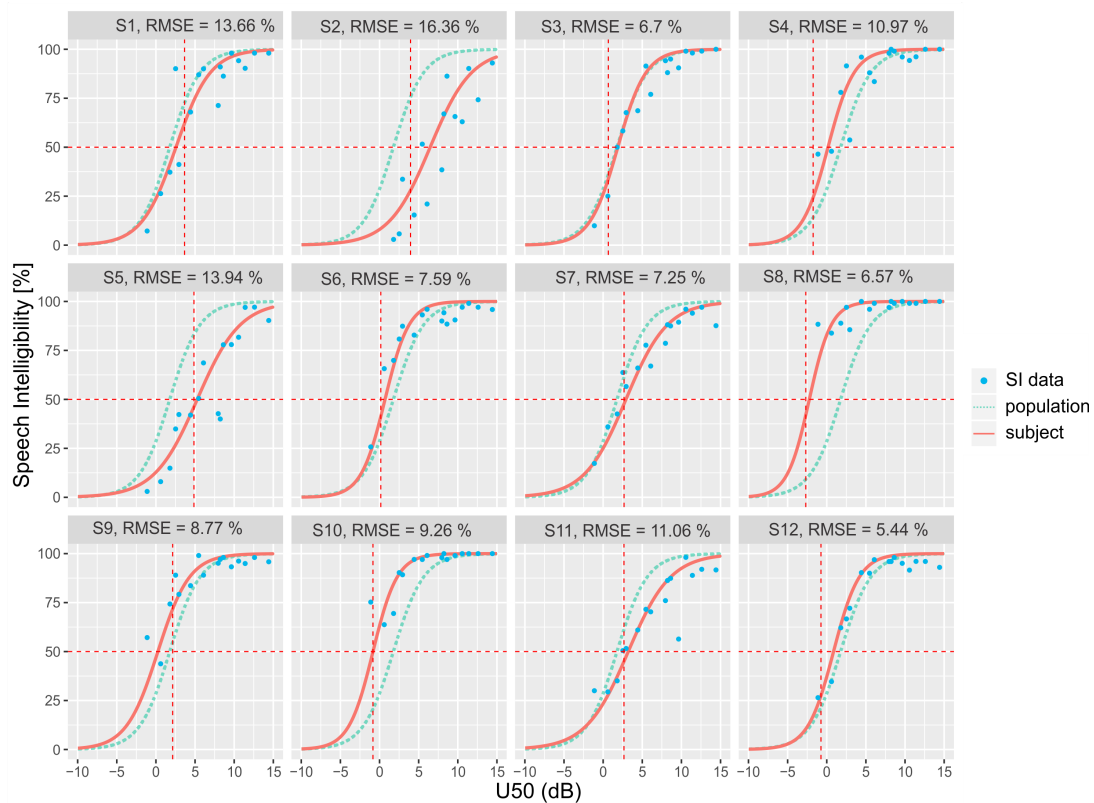


FIGURE 3.2: Individual SI scores measured in the 17 quiet conditions (filled circles), psychometric (sigmoid) functions fitted to the subject's scores (solid lines) and population-level prediction (dashed lines). The vertical dashed lines indicate the measured SRT_N (at the BTE microphone) obtained by each individual.

The corresponding RMS errors are shown in the subject's panel. From Fig. 3.2 it can be seen that the location of the individual predicted \overline{SRT} in quiet, which corresponds to the U50 value of the individual psychometric function at 50% SI, varies substantially across subjects. RMS errors are rather small and vary from 5.44% to 16.36%, with the mean RMS error being 9.8%. The close agreement between the SI scores and the individual fitting curves may be seen as a verification of the U50 as a suitable speech intelligibility predictor. This, in turn, suggests that the early reflections in addition to the direct sound component are beneficial to SI, that LRs are detrimental, and that SI can be expressed as a function of the ratio between DSER and LR.

In order to analyze the variability in SI performance between subjects, the estimated \overline{SRT} as well as the slope k of the fitted psychometric function in quiet conditions are summarized in Table 3.2. The values shown in the table correspond to k_i and \overline{SRT}_{ij} for the i 'th subject and the j 'th category corresponding to *quiet* (see Sec. 3.A), but for simplicity, they are referred to here as slope k and \overline{SRT} .

TABLE 3.2: Slope k and \overline{SRT} in quiet obtained for each subject.

ID	Fitting parameters	
	Slope k (%/dB)	\overline{SRT} (dB)
1	11.4	2.6
2	9.4	6.5
3	13.9	1.9
4	14.9	0.2
5	9.1	5.3
6	16.2	0.7
7	9.4	2.9
8	17.8	-2.2
9	12.3	0.3
10	9.3	-0.8
11	16.1	2.6
12	15.1	0.9

Two observations can be made from Table 3.2. First, the estimated \overline{SRT} in quiet spans a range of 8.7 dB. This finding extends the large inter-subject variability of speech-in-noise performance reported in the literature to the performance in quiet, reverberant conditions. Second, the subjects' slopes were positively correlated ($R^2 = 0.85$) with the subjects' noise presentation level (i.e., the slopes increased with increasing noise level), which varied from 52 dB to 59 dB across subjects and, with the constant speech level of 60 dB SPL (Sec. 3.2.1), resulted in broadband SNRs ranging from 1 dB to 8 dB. In order to rule out the possibility of this correlation being, partially or entirely, a consequence of having constrained the slope k to be fixed for all categories, the correlation was confirmed by a more complete statistical model

whereby the slope k was allowed to vary across categories. Since the resulting correlation was only slightly decreased ($R^2 = 0.83$), the significant correlation was not a result of the constrained slope.

3.3.2 Effect of noise

Because the characteristics of the noise signals differ across rooms due to the different levels of reverberation, the fitting of the psychometric (sigmoid) function to the SI scores in noise was conducted in a room-dependent basis, providing a room-dependent \overline{SRT} of the fitting function but a slope that was fixed across rooms (see Sec. 3.2.5). Each panel of Fig. 3.3 shows the psychometric function that corresponds to the population-level prediction of the measured SI scores in each of the six rooms as a function of U50, including the quiet condition for reference purposes. Shaded areas indicate the population prediction intervals, obtained as described in Sec. 3.2.5. The dashed lines indicate the population-level prediction of the \overline{SRT} . The categories are ordered in Fig. 3.3 according to their predicted \overline{SRT} value. As can be observed, all the categories have a different \overline{SRT} , with the anechoic condition being the most challenging (i.e., the \overline{SRT} is the highest) and the small reverberant room being the least challenging (i.e., the \overline{SRT} is the lowest).

The estimated shift from the estimated \overline{SRT} in quiet that is caused by the effect of noise (i.e., $\overline{SRT}_Q - \overline{SRT}_N$) is shown in Fig. 3.4 as a function of the normalized modulation power spectrum evaluated at 3.1 Hz, the most prominent modulation frequency of the applied BKB-like sentences, i.e., the anechoic target speech (see Sec. 3.2.4). Figure 3.4 reveals that, although the confidence intervals (given by the error bars) are quite wide, there is a strong correlation between the predicted shift of \overline{SRT} obtained in each room and the modulation of the noise. Note that, because the figure is presented as a shift from the \overline{SRT} in quiet, the plot provides information about the effect of noise and is subject independent.

Because the confidence intervals shown in Fig. 3.4 are rather wide, an alternative way to evaluate the importance of accounting for the modulation of the noise consists of comparing the fitting error of statistical models that do and do not account for the modulation of the noise. To allow such a comparison, it is assumed here that the noise-dependent shift parameter (i.e., the \overline{SRT}) of the psychometric function contained in the statistical model described in section 3.2.5 is solely reflecting the effect of the modulations of the noise on SI, and does not simply increase the number of free fitting parameters. Given the strong correlation between this shift parameter and the modulation spectrum of the noise shown in Fig. 3.4, this seems a reasonable assumption. In order to derive a model that does not take into account the modulations of the noise, a simpler version of the original model described in section 3.5 is applied, in which all the seven categories (i.e., the six noisy rooms plus the quiet condition) were treated as a single category (i.e., the model had a single common fixed effect for the slope k as well as for the \overline{SRT}). The mean RMS errors for the original model

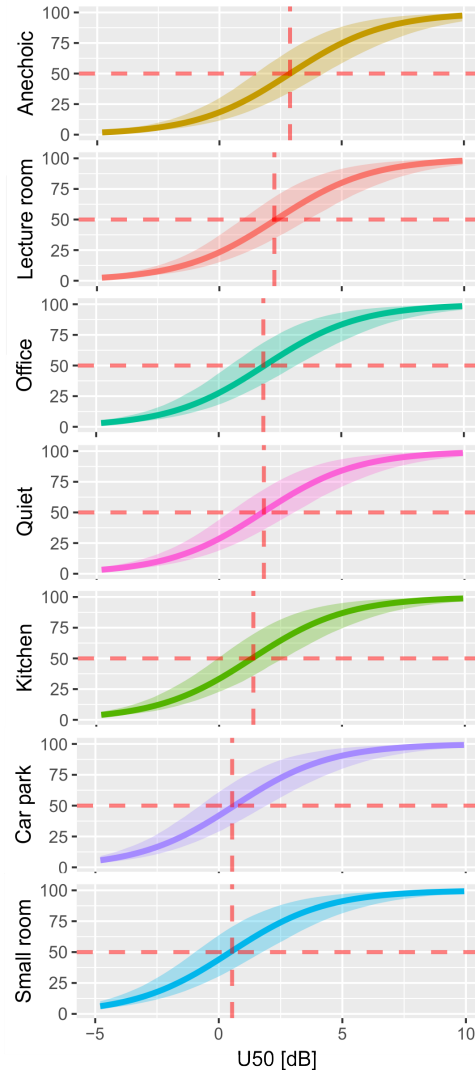


FIGURE 3.3: Group-level SI predictions as a function of the U50 in the six different noisy environments as well as in quiet conditions. The dashed lines indicate the predicted SRT , determined as the U50 at which the predicted psychometric functions cross 50% SI. The SRT closest to the one obtained in quiet conditions (1.77 dB) corresponds to the open-plan office. The shaded area represents the prediction intervals, obtained as described in Sec.[3.2.5]

were 9.8% (standard deviation 3.44%) in quiet conditions and 11.03% (standard deviation 3.13%) in noise. The mean RMS error of the simpler statistical model was 13.46% (standard deviation 3.99%). Hence, the highest errors were obtained with the more basic model, and were consistently lower across subjects than the RMS errors obtained with the original model ($R^2 = 0.82$, see Sec. 3.B for individual RMS errors). This consistent change in RMS error between models further supports the assumption that the modulations of the noise have an impact on SI in rooms.

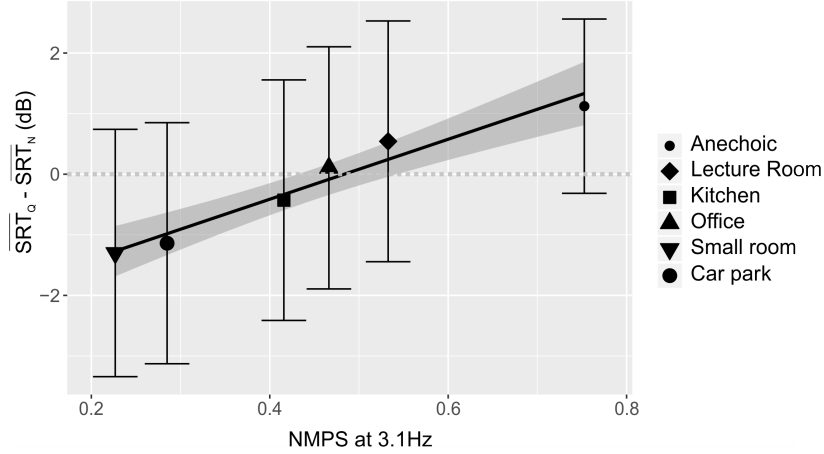


FIGURE 3.4: Shift of the estimated \overline{SRT} between quiet and noise conditions for each room (with 95% confidence intervals) as a function of the normalized modulation power spectrum (NMPS) of the noise signals evaluated at the most prominent modulation frequency of BKB-like sentences. The solid line indicates the regression line ($R^2 = 0.96$) with 95% confidence intervals (shaded gray area)

3.3.3 Relative effects of noise and reverberation

The vertical dashed lines in each of the panels of Fig. 3.2 indicate the SRT_N measured for each subject in the workplace kitchen noise with anechoic target speech. The measured individual SRT_N are very similar to the predicted individual \overline{SRT} in quiet, which are shown in Fig. 3.2 as the U50 value of the fitted individual psychometric function at 50% SI (i.e., the crossing point with the horizontal dashed lines). As the target speech used during the SRT_N measurements was anechoic, and the fitting curves consider only quiet, reverberant conditions, the close agreement between the subjects' individual SRT_N and their \overline{SRT} in quiet suggests that the power of the late reflections of the target speech has the same (detrimental) effect on SI as the power of the noise. This observation supports the assumption, inherent in the U50 metric, that the SI in rooms is limited by the simple addition of the power of the noise (N) and the late reflections (LT), as shown in the denominator of Eq. 3.1.

The results of this comparison are further evaluated in Figure 3.5, where the estimated \overline{SRT} in quiet reverberant conditions is shown as a function of the SRT_N measured in the workplace kitchen noise for anechoic target speech. A linear regression analysis revealed a strong correlation ($R^2 = 0.77$) between the predicted \overline{SRT} in quiet and the measured SRT_N with the regression line shown in Fig. 3.5 by the solid black line and given by $\overline{SRT} = 0.9 \cdot SRT_N + 0.67$. For comparison purposes, the gray dashed line shown in Fig. 3.5 ($\overline{SRT}_Q = \overline{SRT}_N + 0.42$) indicates the relationship between the estimated SRTs in quiet (\overline{SRT}_Q) vs noise (\overline{SRT}_N) as obtained from the U50 predictions for the specific case of the workplace kitchen noise in which the actual SRT_N were measured. The fact that the intercept of the latter curve equals 0.42 (and not 0) is explained by the temporal modulations inherent in the workplace kitchen

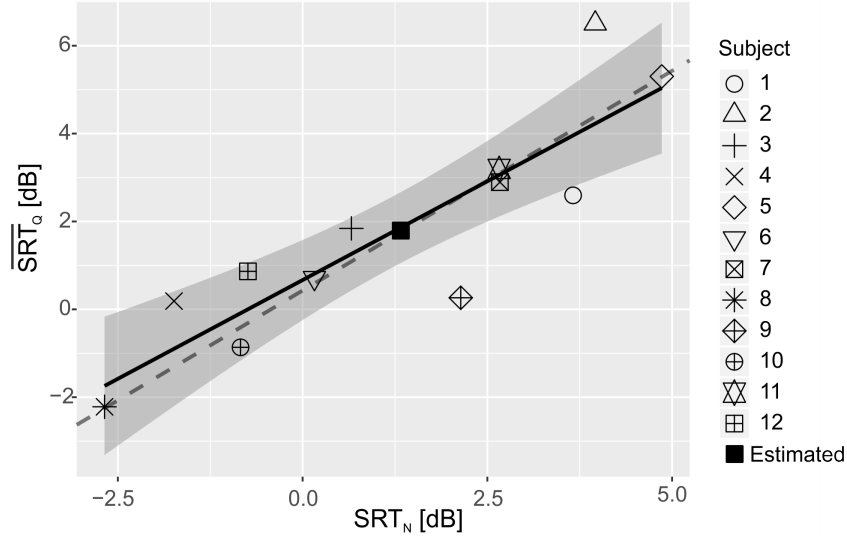


FIGURE 3.5: Estimated \overline{SRT} obtained from the fitting in quiet as a function of the SRT_N measured in noise. The solid regression line depicts the correlation between the two measures. For comparison purposes, the gray dashed line indicates the relationship between the two SRT as predicted by the U50 for the specific case of the workplace kitchen noise that was used for the measurement of SRT_N . The black square is the population-level prediction of the two SRT for this case. See text for more details.

noise, as explained in Sec. 3.3.2 and plotted in Fig. 3.4. The black square shown in Fig. 3.5 refers to the population-level estimates of the \overline{SRT} in quiet and noise. As seen in the figure, the two straight lines (solid and dashed) are close together, with a slightly larger deviation towards lower values of SRT_N . However, throughout the entire SRT_N range, the dashed line falls within the 95% confidence intervals of the linear fit. The similarity of these two straight lines further supports the conclusion suggested above that the late reflections of the target speech have the same detrimental effect on SI as noise.

3.4 Discussion

The results presented in this study suggest that the U50 is a measure suitable for predicting the effect of noise and reverberation on SI scores. However, results in noisy conditions show that the modulations contained in the noise signal have an effect on SI, which is not accounted for in the current formulation of the U50.

3.4.1 Speech intelligibility in quiet

As shown in Sec. 3.3.1, the U50 is able to predict speech intelligibility performance of CI recipients obtained in quiet rather accurately for a large range of reverberant conditions. This suggests that the U50 could potentially be used in the evaluation of whether a given venue is amenable to CI recipients. Table 3.3 shows the correspondence between the U50 and expected SI scores obtained with NH listeners and

CI users. In order to account for the large variation in the SI performance observed across CI recipients, the expected scores are shown for three types of CI recipients: CI-P (population-level predicted SI), CI-L (SI obtained with the person with the most difficulties or lowest scores), and CI-H (SI obtained with the person with the least difficulties or highest scores). Speech intelligibility scores for NH listeners shown in Table 3.3 were obtained from the curves shown in IEC (2003). From table 3 it can be deduced that all CI recipients need far higher U50 values than NH listeners to achieve similar SI, which is more pronounced towards lower quality categories. Even though the best performing CI recipient (CI-H) shows a SI performance that is almost as good as for a NH listener for the U50 values that are rated as "excellent" by NH listeners (i.e., with scores around ceiling), the difference in performance between them increases rapidly with decreasing quality category. This decay in performance with decreasing quality category is even further pronounced for the poorer performing listeners, with some showing clearly below 100% SI even under "excellent" conditions and near 0% intelligibility under conditions that are rated as "fair" by NH listeners. However, it should be noted here that the comparison between the SI of NH listeners and CI users must be taken with care, as they both have been obtained with different methodologies, speech material and even scoring methods.

Even though SI with CIs in quiet can be accurately predicted with the U50, subject variability and the fact that overall scores are much lower than with NH individuals (in particular at lower U50 values) prevent room-acoustic solutions from being the most practical option to ensure a satisfactory SI in venues. As shown in Sec. 3.3.3, the predicted \overline{SRT} in quiet varies from -2.2 dB to 6.5 dB between subjects, with a population-level prediction of 1.77 dB. Hence, room acoustic requirements based on population-level predictions of CI recipients would have the risk of not being suitable for a large proportion of their population, as these predictions are representative of only a small part of it. Moreover, population-level predictions of SI are much lower with CI users than with NH listeners. Rooms falling within a specific rating category for NH listeners fall within one or two rating levels below for CI users. Hence, it is rather unlikely that acousticians and architects will design venues to satisfy the requirements for CI recipients. Instead, the provided quality ratings for CI recipients may be rather used as a guide for alternative solutions that can be installed inside a venue, such as hearing loops, FM systems, or Bluetooth systems, which can send the speech signal of a talker directly to the speech processor of the CI recipients. Additionally, the quality ratings in combination with the U50 measure may be used to control or improve (customized) speech enhancement methods such as de-reverberation algorithms, directional microphones, or scene classifiers with the CI's speech processor. In this regard, however, it should be noted that SI scores reported in the present study were obtained with omnidirectional BTE microphone directionality and most (adaptive) signal enhancement features were turned off. Hence, it can be speculated that by turning on these features, overall SI would be improved and the difference to NH listeners reduced. However, these and other limitations of the

present study are out of the scope of the present study and are further discussed in Sec. 3.5.

TABLE 3.3: U50 values, equivalent STI values and quality ratings according to IEC (2003), and expected SI with NH listeners. Expected SI for CI recipients are shown in three different cases: subject with the most difficulties (CI-L), population-level predictions (CI-P), and subject with the least difficulties (CI-H)

NH listeners				CI recipients: SI (%)		
U50 (dB)	STI	Rating	SI (%)	CI-L	CI-P	CI-H
$-18.5 < \text{U50} < -8.5$	0 - 0.3	Bad	0 - 15	0	0	0 - 1
$-8.5 < \text{U50} < -3.5$	0.3 - 0.45	Poor	15 - 70	0 - 2	0 - 6	1 - 29
$-3.5 < \text{U50} < 1.5$	0.45 - 0.6	Fair	70 - 98	2 - 13	6 - 46	29 - 93
$1.5 < \text{U50} < 6.5$	0.6 - 0.75	Good	98 - 100	13 - 50	46 - 92	93 - 100
$6.5 < \text{U50} < 15$	0.75 - 1	Excellent	100	50 - 96	92 - 100	100

3.4.1.1 Room parameters relevant to SI in quiet

One of the benefits of using the U50 as a predictor of SI is the fact that it is easily interpretable in terms of room acoustics, as it can be broken down into a set of descriptive room parameters. By using a statistical room model, the U50 in quiet can be expressed as:

$$\text{U50}|_{\text{quiet}} = 10 \cdot \log \left(\frac{1 - e^{-0.69/RT} + 10^{DRR/10}}{e^{-0.69/RT}} \right) \quad (3.4)$$

where RT is the reverberation time and DRR is the direct-to-reverberant ratio. Equation (3.4) was obtained from Bistafa and Bradley (2000), where the term associated with the theoretical direct-to-reverberant ratio was here expressed as a function of the actual, measured DRR, thereby avoiding any assumptions of source-receiver directivity. Under diffuse sound field assumptions, and for an omnidirectional sound source, the DRR can be expressed as:

$$\text{DRR} = 10 \cdot \log \left(\frac{0.16V}{16\pi r^2 RT} \right), \quad (3.5)$$

where V is the volume of the room and r expresses the talker-to-listener distance. Combining Eq. (3.4) and Eq. (3.5) leads to:

$$\text{U50}|_{\text{quiet}} = 10 \cdot \log \left(\frac{1 - e^{-0.69/RT} + \left(\frac{0.16V}{16\pi r^2 RT} \right)}{e^{-0.69/RT}} \right) \quad (3.6)$$

Equation (3.6) is a signal-to-noise measure where the term $\left(\frac{0.16V}{16\pi r^2 RT} \right)$ expresses the direct sound contribution, the term $(1 - e^{-0.69/RT})$ expresses the contribution

of early reflections, and the term $e^{-0.69/RT}$ found in the denominator expresses the detrimental effect of the late reflections. Hence, the U50 is simply described by the volume and reverberation time of the room as well as the talker-to-listener distance.

At a fixed distance, Eq. (3.6) suggests that:

- (1) The contribution of the direct sound component depends on the RT and the volume of the room. The weakest contribution is found in small volumes and high RTs. Rooms with longer RTs may have stronger direct sound contributions than rooms with shorter RTs. For example, the DRR at 1.3m is 10.3 dB higher in the car park than in the small reflective room (see Table 3.1). This occurs because the volume of the car park is much larger than that of the small reflective room. Rooms with longer RTs may also have weaker direct sound contributions than rooms with shorter RTs. For example, the DRR at 1.3m is 4.26 dB lower in the workplace kitchen than in the lecture room. This occurs because the volume of the two rooms is the same while the RT in the workplace kitchen is 0.68s and in the lecture room is 0.46s. More generally, Eq. (3.5) suggests that a room with volume V_1 and RT_1 may have a higher or lower direct sound contribution than another room with volume V_2 and RT_2 if the ratio V_1/RT_1 is higher or lower than V_2/RT_2 respectively.
- (2) The contribution of the direct sound component on the U50 prevents the RT from being a good indicator of SI. Rooms with longer RTs and stronger direct sound contributions do not necessarily present lower U50 values. This is, for example, the case for the car park, which has a longer RT than for example the small reflective room but provides better SI. This occurs because the decrease of the term associated with the contribution of the early reflections is counteracted by the increase of the term associated with the direct sound contribution. Hence, it is not necessarily true that lower SI will be obtained for longer RTs. In contrast, rooms with longer RTs and weaker direct sound contributions unequivocally present lower U50s and hence, worse SI. This effect may for instance be observed when different reverberant conditions are obtained by varying the absorption material of a single room. For example, the workplace kitchen and the lecture room have the same volume, but the workplace kitchen has a far longer reverberation time and incurs significantly lower intelligibility scores.

In real-life situations, very long RTs are commonly encountered in large rooms, where the contribution of the direct sound at conversational distances can be quite significant. In fact, among the rooms employed here, the highest DRRs are observed in the room with the longest RTs (see Table 3.1). Several studies that have evaluated the effect of reverberation on SI with CIs have obtained high RTs (e.g. in the order of one second) by altering the absorption material of a small room (e.g. 76.8 m³ Kokkinakis, Hazrati, and Loizou, 2011; Hu and Kokkinakis, 2014). In these cases, the

contribution of the direct sound component, which is already weak given the dimensions of the room, decreases with increasing RT, as the room volume is kept fixed. This leads to rather low U50 values that decrease monotonically and quite rapidly with increasing RT. Hence, it is not surprising that these studies found such a strong effect of reverberation on SI with CIs. In situations like these, it is advantageous to report the STI or the C50, as their values can be more easily generalized to other rooms and talker-to-listener distances. However, in many of these studies, likely because SI was uniquely explained by the RT, the reverberant conditions have been described by just the RT. As a consequence, the results found in these studies cannot be easily extrapolated to real life, even though rooms with similar acoustic properties may well exist (e.g., an empty living room).

3.4.1.2 Additional room acoustic factors

Studies that have obtained different reverberant conditions by altering the absorption of small rooms have potentially altered significantly the frequency response of the signal across reverberant conditions. This may occur for several reasons. First, because small reverberant rooms usually have boosted high frequencies, and because high frequencies are more easily absorbed, the low-pass filtering effect of adding absorption material may have a significant impact on the frequency response. Second, because the relative level of strong single early reflections, whose colouration effect may be significant, decrease with increasing absorption material, the frequency response may potentially become more balanced with increasing absorption. Third, because the bandwidth of room modes increases with increasing absorption material, the Schroeder frequency (the frequency below which the sound field is dominated by isolated modes) increases with increasing RT (Schröder, 1954) in such a way that modes in one condition (e.g., source or receiver location) may not be present in a different condition. As a result of all these factors, the frequency response may be different across reverberant conditions.

While the frequency response may not have a strong impact on SI with NH listeners in quiet, it is unclear whether this is the case for CI users. The N-of-M algorithm implemented in current CI speech processors selects the frequency bands with higher envelope amplitudes. If a frequency band at the listener position has prominent peaks arising from any of the aforementioned factors or a combination thereof, that band may be constantly selected, even though it may not carry any important speech information. As different reverberation conditions may have different frequency responses, the selected electrodes will consequently be different across conditions. In particular, given how the energy of high frequencies is progressively absorbed for decreasing RTs, selected electrodes may progressively be shifted towards low frequencies as more absorption material is added.

Previous studies have overlooked this potential confounding factor (e.g. Kokkinakis, Hazrati, and Loizou, 2011; Hu and Kokkinakis, 2014). For example, in Hu and Kokkinakis (2014), anechoic speech was convolved with the early reflections (plus

the direct sound component) of different impulse responses obtained by varying the absorption material of a single room with a volume of 76.8m^3 . Because the SI data showed that for higher RTs SI decreased, the authors concluded that CI users exhibit a reduced ability to fuse the direct sound with early reflections. However, another potential explanation is that the early-reflections do not contribute equally to intelligibility across conditions due to the different frequency responses. Alternatively, even if early reflections contribute equally to intelligibility across conditions, if they are not as beneficial as the direct sound component (e.g., Arweiler and Buchholz, 2011), the fact that their relative contributions differ across conditions (because the relative direct sound contribution is reduced for increasing RTs) may explain the differences in speech intelligibility observed. Although it is unclear how relevant the frequency response of small rooms is to SI with CIs, SI scores obtained in single small rooms must be taken with care, as they may be very specific to the room, RT and even target and listener position.

3.4.2 Speech intelligibility in noise

In Sec. 3.3.2 it was shown that even though SI performance of CI recipients can be reasonably well predicted by a basic model that does not include any room-dependent fitting parameters, providing such parameters clearly improved the accuracy of the model predictions in noise, as measured by the RMS error.

The results presented in section 3.3.2 and in particular Fig. 3.4 revealed that the modulation of the noise is a parameter relevant to SI. In general, reverberation flattened the temporal envelope of the modulated noise and thereby reduced the relative power of the noise modulations that are most detrimental for understanding speech. Given the different levels of reverberation in the different rooms, their effect on the temporal noise envelope and thus, the modulation spectrum, varied also across rooms.

The data suggest that non-modulated noises shift the fitted psychometric function to the left (making it easier to understand speech) and modulated noises shift the fitted psychometric function to the right (making it harder to understand speech). This is consistent with the findings reported in Chapter 2, where the anechoic noise (i.e., the most modulated noise) was found to be the most detrimental noise. Nevertheless, in that study, a quantitative analysis of the effect of the noise modulation was not reported, as the reverberation of the target speech was different across rooms. In fact, only anechoic noise stood up as different from the rest in terms of SI. As opposed to Chapter 2, the present study enables a clearer comparison between noise types, as the effect of the target reverberation is accounted for in the U50.

Although the U50 does not account for the modulation of the noise, it seems fairly straightforward to include a term to account for it. For this, a modulation-dependent SRT shift could be introduced to the U50 measure, which could be expressed as a function of the normalized modulation power spectrum of the noise using the regression line obtained in Fig. 3.4 as a mapping function. Nevertheless, the final

expression will depend on the way in which the modulation spectrum of the noise is quantified. In the present study, the modulation of the noise was reduced to a single value at the most prominent modulation frequency of the (anechoic) target speech, which was sufficient to infer an effect of noise modulation on SI. However, a more complete description of both the noise and the target signal in the modulation domain will most likely lead to more accurate variations of the U50, which is beyond the scope of this study. Future work in this direction should probably consider models where the SNR in the modulation domain is estimated in a more exhaustive way (e.g. Jørgensen and Dau, 2011).

Given the great dependence of SI with CIs on the noise and target modulation characteristics as well as on their interaction, it seems reasonable to suggest that future work should consider SI predictions techniques that are based on the modulation domain. For example, the model presented in Jørgensen and Dau (2011) could potentially be adapted to include the possibility to predict SI in quiet conditions in a similar way as *ModA* does (Chen, Hazrati, and Loizou, 2013). Further improvements could consider the signals at the output of the speech processor (as in Watkins, Swanson, and Suaning, 2018) and even incorporate subject-dependent fitting parameters such as the N-of-M algorithm or the electrical dynamic range, as determined by the T and C levels.

3.4.3 Speech Transmission Index (STI)

The strong relationship between the STI and the U50 is usually described by means of a regression line that acts as a mapping function between the two metrics (e.g. Bistafa and Bradley, 2000; Nijs and Rychtáriková, 2011). However, the actual relationship between them is somewhat more complex than a straight line and it can be shown to depend on the RT and the DRR (see Sec. 3.C for more details). This section is devoted to investigate whether the subtle deviations between the STI and the U50 are relevant to the prediction of SI with CIs. To conduct this comparison, the STI was calculated as per IEC (2003), that is, based on the impulse response and including auditory masking. However, in order to conduct a fairer comparison with the present U50 formulation, the STI calculation was modified so that the actual frequency response of target and noise signals (in octave bands) was accounted for and only the female STI was considered, as the target speech was produced by a female talker. Following the same procedure as for the sigmoid-based fitting of the U50 (i.e. using the same statistical model), the STI was fitted separately to the SI data in noise (i.e., treating each room separately) and in quiet (Sec. 3.2.5). Following the steps introduced in Sec. 3.3.1, mean RMS errors with the STI-based predictions were 9.7% in quiet (standard deviation 3.48%) and 11.02% in noise (standard deviation 3.02%). For the case that a more basic fitting model was applied that did not allow for any noise specific effects, the RMS error in noise increased to 14.62% (standard deviation 3.99%). Following the same order, the RMS errors obtained with the U50 were 9.8%, 11.03% and 13.46%.

Hence, in all cases, RMS errors were comparable between the two methods even though small differences could be observed for individual conditions.

3.5 Outlook and limitations

The speech intelligibility data considered in this study was obtained in unilateral CI recipients with most adaptive features turned off and using only the front BTE microphone of the speech processor with an omnidirectional directivity pattern. Future work should consider the benefits of fixed microphone directivity, adaptive beamformers, and other signal enhancement features available in CIs. For the case of a fixed directional microphone, it is expected that the U50 is able to correctly predict the provided benefit in SI. However, for any other, in particular adaptive, speech enhancement feature it is questionable if the U50 can correctly predict the provided benefit. In these cases, the current U50 implementation will need to be extended, for instance, by applying a short-term frequency analysis or by considering the SNR in the modulation domain. Although ADRO can assist resolving issues related to audibility, studies have shown no SI benefit in quiet reverberant conditions (Ali et al., 2014). In the present study, audibility was controlled by fixing the target speech level to 60 dB SPL. However, in realistic conditions at far target-to-listener distances, audibility may become important due to the limited acoustic power that can be provided (or sustained) by a talker and the rather rapid decay of the direct sound level with increasing distance.

Another important factor that has not been considered in the present study is the effect of a second implant. At least in spatially asymmetric noise conditions, CI recipients can receive a significant benefit from a second implant, which is mostly due to better-ear listening. Due to the spatial separation between the target speech and an interferer, the SNR at one ear can be significantly better than at the other ear due to head shadow effects. Having access to the implant with the better SNR can then provide a substantial benefit. However, this spatial benefit typically decreases with increasing number of interferer as well as with reverberation (e.g., Peissig and Kollmeier, 2002; Lavandier and Culling, 2008). The benefit provided by a second implant in quiet reverberant conditions is unclear. For target speech from the front, even NH listeners may receive only a small benefit from having access to two ears (e.g., Westermann and Buchholz, 2015). This might be due to the fact that late reverberation is rather diffuse and will therefore have the same effect on both ears. However, this may be different for the benefit provided by early reflections (e.g., Arweiler and Buchholz, 2011), which, in turn, will depend on the given room as well as the location of the receiver and the source. However, more research is required to reveal the true benefit in SI that is provided by a second CI in reverberant conditions.

Finally, although each subject was tested at a different noise level (or SNR) to avoid ceiling and floor effects, the SI data considered in this study was measured at a single target and noise level for each individual subject. Testing subjects at different

SNRs may, in particular, have an effect on the slope of the predicted psychometric function. As seen in Sec. 3.3.1, the subjects' slopes were positively correlated with the presentation level of the noise. This is important here because the non-linear mixed-effects model that was applied to predict the individual speech intelligibility data included two free fitting parameters per subject, the slope as well as the shift (or \overline{SRT}) of the applied psychometric (model) function. As a consequence, the subject specific ability was confounded by the tested SNR. Hence, it is unclear if the differences in slopes across subjects are due to their individual performance or due to the tested SNR. Segregating these two effects will help to separate the effects that are purely signal-driven, and may therefore be predicted by the U50, from effects that are subject specific, and cannot be predicted by the U50. Hence, future research will need to test SI in CI recipients at different SNRs (or U50 values) and then compare the performance of subjects at the same SNRs. Of course, such comparison will be complicated in view of the large inter-subject variability in SI performance of CI recipients and the associated floor and ceiling effects.

3.6 Conclusions

U50 values were fitted to speech intelligibility data of 12 cochlear implant users tested unilaterally under a large number of realistic reverberant conditions. The performance of the U50 as a room-based SI predictor was shown to be highly accurate in quiet conditions. Moreover, by applying a simplified (statistical) model of the room impulse response within the U50 formulation, it was shown that the discrepancies observed in previous studies on the effect of reverberation on SI in CI recipients in quiet was largely due to the fact that some studies applied rather small reverberant rooms, which provided overly low U50 values that resulted in an overly low SI. In noisy conditions, the accuracy of the U50 (as measured by the RMS error) was slightly reduced. This was largely due to reverberation smoothing the envelope of the applied modulated noise, a factor that improved speech intelligibility but is not captured by the U50. By applying a basic modulation spectrum analysis, it was shown that the modulation power of the noise, at the modulation frequencies most critical for speech, provides a good estimator for the benefit in SI provided by the smoothing effect of the noise envelope due to reverberation. This suggests that future formulations of the U50 (or alternative SI models) should take into account the modulation spectrum of the noise (and the target speech). This study also showed that the same conclusions drawn for the U50 in both quiet and noisy reverberant conditions can be applied to the speech transmission index.

Appendices

3.A Nonlinear mixed-effects model

The non-linear mixed effects model was given by:

$$\begin{bmatrix} k_i \\ \overline{SRT}_{ij} \end{bmatrix} = A_j \begin{bmatrix} k \\ \overline{SRT}|_{j=1} \\ \overline{SRT}|_{j=2} \\ \overline{SRT}|_{j=3} \\ \overline{SRT}|_{j=4} \\ \overline{SRT}|_{j=5} \\ \overline{SRT}|_{j=6} \\ \overline{SRT}|_{j=7} \end{bmatrix} + B \begin{bmatrix} k_{1i} \\ \overline{SRT}_i \end{bmatrix} = \beta_j + b_i \quad (3.7)$$

$$b_i \sim N(0, \Psi), \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

where i denotes the subject, j the category, k_i represents the slope of the curve (in %/dB) at 50% SI, \overline{SRT}_{ij} is the U50 at 50% SI, Ψ is the variance–covariance matrix, B is the 2x2 identity matrix and A_j depends on the category. For example, for room 4 in noise A_j is:

$$A_4 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

3.B Individual RMS errors

Table 3.B.1 shows the RMS errors obtained during the U50 fitting for every subject. The second ($U50|_{quiet}$) and fourth ($U50|_{noise}$) columns show the RMS error obtained with the model described in Sec. 3.2 for the quiet and noise conditions respectively. The third column (Basic $U50|_{noise}$) shows the error for the noise conditions of a more basic statistical model in which the fitting was room-independent.

TABLE 3.B.1: RMS error obtained for each subject in three different cases: U50 in quiet conditions ($U50|_{quiet}$), room-independent U50 in noise conditions (Basic $U50|_{noise}$) and U50 in noise conditions ($U50|_{noise}$). The second and fourth column correspond to the errors obtained with the category-dependent model described in Sec. 3.2.

ID	RMS error (%)		
	$U50 _{quiet}$	Basic $U50 _{noise}$	$U50 _{noise}$
1	13.66	10.8	8.54
2	16.36	23.64	20.06
3	6.7	15.44	10.25
4	10.97	14.22	10.79
5	13.94	10.62	9.41
6	7.59	17.05	13.49
7	7.25	12.57	11.33
8	6.57	11.93	9.01
9	8.77	10.18	9.26
10	9.26	9.38	9.64
11	11.06	10.90	10.63
12	5.44	14.81	9.93
Mean	9.8	13.46	11.03

3.C Relationship between the C50 and the STI

In order to analyze further the relationship between the STI and the C50, this section compares the two metrics by means of simulated and measured RIRs. The RIR model consists of a delta function expressing the direct sound immediately followed by the reverberant part of the signal, which decays exponentially over time t as $e^{\frac{-13.8t}{RT}}$. The calculation of the C50 from this RIR model leads to Eq. (3.4). As for the STI, the first step is to calculate the modulation transfer function, which can be expressed as (Houtgast, Steeneken, and Plomp, 1980):

$$m(F) = \frac{(A^2 + B^2)^{1/2}}{C} \quad (3.8)$$

where

$$\begin{aligned}
A &= 1 + \frac{r^2}{r_c^2} \left[1 + \left(\frac{2\pi F \cdot RT}{13.8} \right)^2 \right]^{-1} \\
B &= \frac{2\pi F \cdot RT}{13.8} \frac{r^2}{r_c^2} \left[1 + \left(\frac{2\pi F \cdot RT}{13.8} \right)^2 \right]^{-1} \\
C &= 1 + \frac{r^2}{r_c^2}
\end{aligned}$$

with r the source-receiver distance and r_c the critical distance. Note that Eq. (3.8) corresponds to Eq. (13) in Houtgast, Steeneken, and Plomp (1980) where the numerator and denominator have been multiplied by r^2 and where the directivity of both source and receiver is assumed to be omnidirectional. In order to make Eq. (3.8) independent of the distance r and the critical distance r_c , the term $\frac{r^2}{r_c^2}$ is replaced by the inverse of the DRR on a linear scale. As with the U50, the theoretical DRR is then here replaced by the measured DRR. This way, no assumptions about the directivity of the system need to be made. Once the modulation transfer function is obtained for all the modulation frequencies, the STI can be easily obtained by following the steps described in equations 3 to 5 (Houtgast, Steeneken, and Plomp, 1980).

In the calculation of the STI and the C50, the RIR is assumed to be frequency-independent. For simplicity, the $[0 - 1]$ clipping involved in the STI calculation is removed, thus allowing it to have values greater than one. The modulation frequency F takes values from 0.63 Hz to 12.6 Hz in third octave bands. As can be seen in Eq. (3.8) and Eq. (3.4), both the C50 and the STI depend on the RT and on the DRR. Hence, the simulations are based on sampling these two parameters independently. As for the actual RIRs, a total of 95 real RIRs obtained in a variety of rooms and distances were included in the analysis. The DRR, the C50, the STI and the RT were all obtained as the mean across octave bands. The calculation of the STI from measured RIRs was reduced to its original version (Houtgast, Steeneken, and Plomp, 1980), where auditory masking is not accounted for.

Figure (3.C.1) shows the results of obtained from both simulated and measured RIRs where RT is used as a parameter and where, for clarity, the DRR dependence is not explicitly shown. The results show a reasonable agreement between the calculations obtained from simulated and measured RIRs. With decreasing C50 values, the STI plateaus from a C50 that depends on the RT. In particular, for lower RTs, the STI plateaus at higher C50 values than for higher RTs. As the only parameter that varies along a given line is the DRR, the results suggest that the STI is less sensitive than the C50 to the contribution of the direct sound. The fact that different RTs may lead to moderate C50 values (e.g. 0 to 5 dB) leads to a higher variance of the C50-STI dependency at this range. The regression line shown in Fig. (3.C.1) has been obtained from Nijs and Rychtáriková (2011) and indicates that, when it is not of interest to account for the effects of DRR and RT, a simple regression line may be a good indicator of the relationship between the two metrics.

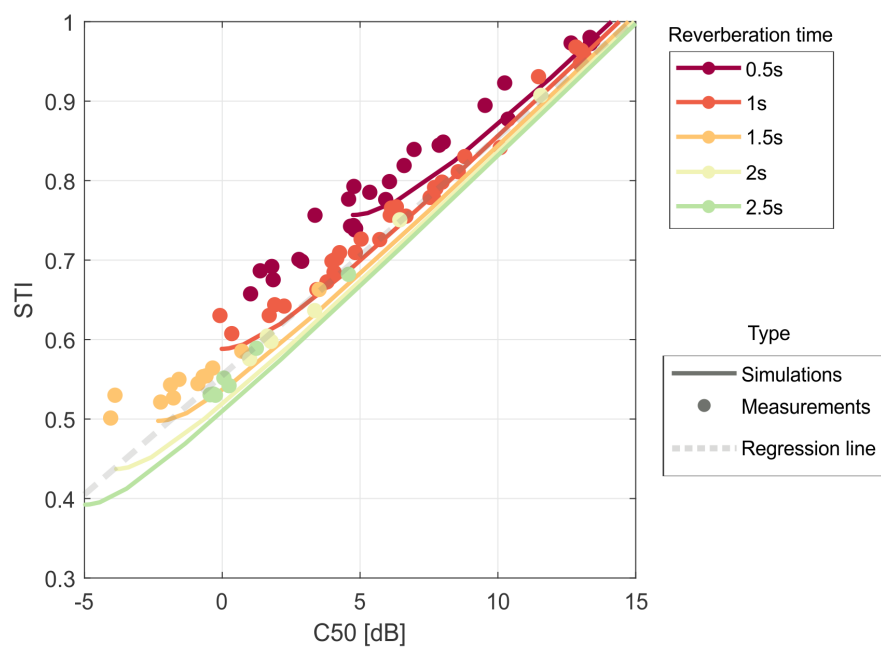


FIGURE 3.C.1: Relationship between the C50 and the STI obtained from simulated RIRs (continuous lines) as well as measured RIRs (solid circles) for different RTs. The reported RTs of measured RIRs have been rounded to the nearest 0.5. The regression line depicts the relationship commonly reported in the literature (in this case, Nijs and Rychtáriková, 2011)

Chapter 4

Effect of test realism on speech-in-noise outcomes in bilateral cochlear implant users

Abstract

Current speech intelligibility tests conducted in the laboratory employ target speech and interfering noise signals that do not resemble what a person encounters in their daily lives. The present study has two main goals. First, to evaluate the effect of laboratory-based test realism on speech intelligibility outcomes of CI users. Speech intelligibility scores of 15 bilateral cochlear implantees under three different test realism levels were compared at two different SNRs. The levels included (1) standard BKB-like sentences with spatially separated standard babble noise, (2) standard BKB-like sentences with three-dimensional recordings of actual situations and (3) a variation of the second realism level where the sentences were obtained from natural effortful conversations. Speech intelligibility was consistently easiest in the most artificial test and hardest in the most realistic test. A low correlation was found between these two levels of realism at the lowest SNR (-2.2 dB), indicating a low consistency of subjects' performance attained between them. The second goal was to conduct an exploratory investigation of speech intelligibility of bilateral cochlear implant users, including bilateral benefit, under more realistic conditions. Speech intelligibility of the more realistic speech material was measured in six different acoustic scenes with SNRs ranging from -5.8 dB to 3.2 dB. Measured scores were in general low, with mean scores around 60% at the highest SNR. Bilateral benefit provided on average a 7% benefit over unilateral speech understanding in the better-performing ear. The findings are discussed alongside potential implications and limitations.

4.1 Introduction

Communication and social connectedness are key to being an active member of the community. However, both hearing impairment and listening in adverse acoustic

environments can impede communication, often resulting in limited social participation and negatively affecting psychosocial outcomes, education, economic stability and independence. However, as reported in a number of studies, speech intelligibility (SI) outcomes measured with typical laboratory tools often do not reflect what people experience in their daily lives (e.g., Cord et al., 2004; Cord et al., 2007).

The mismatch between outcomes in the laboratory and those in the real world is often attributed to the fact that relevant variables such as dynamic variations in space and level, realistic reverberation, or intelligible competing conversations are not included in current speech-in-noise tests (Best et al., 2015). For example, communicating with friends and family often takes place in unideal spaces such as parks, community centres, cafés and restaurants, where a wide variety of dynamic noises arrive from unpredictable directions at unpredictable times. More work is therefore required to understand how these types of environments - spaces in which people occupy and communicate in their everyday lives - affect SI of individuals who use CIs.

Alternative methodologies aimed to obtain more ecologically valid SI outcomes have previously been suggested in the literature. Some of these are based on the *binaural* reproduction of real acoustic environments (e.g., Killion et al., 1998; Culling, 2016) while others are based on multi-channel loudspeaker systems (e.g., Gifford and Revit, 2011; Compton-Conley et al., 2008; Favrot and Buchholz, 2010; Best et al., 2015). While laboratory-based realistic testing based on multi-channel loudspeaker systems has multiple benefits in terms of control and replicability, it remains unclear whether it is worth the investment to set up such complicated testing facilities. Motivated by this question, Best et al. (2015) measured speech reception thresholds (SRTs) under two different levels of realism. One of the levels corresponded to a "standard" anechoic babble noise while the other level corresponded to a more realistic environment where a similar noise source layout was simulated in a café. They found that the SRTs in normal-hearing listeners were similar in the two environments but for hearing-impaired listeners increased much faster with increasing hearing loss in the realistic than in the standard environment. Although this demonstrated an important effect of increasing test realism, the SRTs in the two environments were highly correlated.

There are multiple dimensions in which clinical and laboratory assessments can be improved to closer equate to listening in the real-world, including types of background noise, speech materials, signal-to-noise ratios (SNR) and the speech perception task itself. Improving ecological validity across one or all of these dimensions may result in a performance measure that matches better what is observed in the real world (Cord et al., 2007; Jerger, 2009) and may provide a greater understanding of the communicative challenges faced by individuals with CIs.

One ecologically relevant aspect that could be improved in current SI tests is in regards to the noise signals. The effect of background noise on SI is currently evaluated using steady noise, temporally or spectrally modulated noise, or babble noise (Jerger,

2009). Added to these highly unrealistic signals is the use of oversimplified experimental setups in which a single loudspeaker is used to reproduce the target speech while two (Rana et al., 2017) or three (Mauger et al., 2014) loudspeakers spatially separated reproduce the noise signals. This gives rise to background noises where real-world attributes like reverberation, frequency response or the spatial characteristics nor the dynamic and unpredictable behaviour of the involved noise sources are adequately reproduced.

Another aspect that could be incorporated into speech-in-noise tests concerns the speech material. Speech materials typically used in both clinically and laboratory-based testing diverge from the speech encountered in the real-world. For example, sentence materials generally comprise short, well-formed and concise sentences that are recorded by a trained speaker carefully reading them aloud in anechoic conditions. While this approach permits greater experimental control, the materials lack the complexity and variation that challenge listeners in the real-world. Moreover, they do not capture the various strategies used by interlocutors to ease communication in background noise such as employing Lombard speech (Lombard, 1911) to “talk above the noise” in order to be understood. During such speech, acoustic properties such as fundamental frequency, vowel duration and spectral information change (Lu and Cooke, 2008) in addition to the overall level.

Speech-in-noise tests could be further improved by considering the SNRs that are most commonly encountered in real life (Pearsons, Bennett, and Fidell, 1976; Smeds, Wolters, and Rung, 2015; Wu et al., 2018; Weisser and Buchholz, 2019). In real-world environments, the levels of target and noise signals are not independent. This occurs because people tend to raise their voices (or move closer to each other) to effectively improve the SNR. Hence, a given background noise with a certain sound pressure level will entail a certain speech level. Current speech-in-noise tests do not include this dependency between target and noise levels, and are usually based on the Speech Reception Threshold (SRTs, e.g., Keidser et al., 2013). Even though the SRT is a tool to effectively avoid ceiling and floor issues that may be encountered in tests with fixed SNRs, they typically seek the SNR at which 50% SI is measured. This SNR does not provide much information about real life experience, as the resulting SNRs are far lower than observed in the real world (Smeds, Wolters, and Rung, 2015) and people would rarely be willing to maintain conversation for long periods of time with such a low level of speech understanding.

The present study incorporates three novel ecologically relevant variables into speech-in-noise tests: more realistic noise signals, target speech signals and SNRs. For the noise signals, a spherical array of microphones was employed to record real-life acoustic scenes including an office, a living room, a small church, a dinner party, a café and a food court (Weisser et al., 2019). The recorded scenes were processed and reproduced by a spherical array of 41 loudspeakers. During the listening tests, participants were seated in the centre of the loudspeaker array, located in an anechoic chamber. A similar procedure was followed to ensure that the speech signals carried

the reverberation of the same environment. For the target speech signals, a newly-developed sentence material was employed, which was extracted from natural conversations recorded at three different vocal effort levels (soft, moderate and raised; Kelly M. Miles et al., 2019, in preparation). As opposed to other "standard" speech materials, the sentences in this case were not always self-contained (e.g., "that's interesting isn't it"), were not carefully articulated and presented a higher variability in speed, pronunciation and other linguistic and communication relevant characteristics. For the SNRs, the present study relied on the findings of a recent study that dealt specifically with the SNRs arising from Lombard speech under different levels of background noise (Weisser and Buchholz, 2019). The SNRs were measured from actual conversations between two people at two different distances while they were presented with a range of noise scenes via highly open headphones. In their study, the authors provide an estimate of realistic SNRs as a function of the noise level, the talker-to-listener distance as well as the gender of the speaker.

The present study applies the above methods to address two main goals. First, to evaluate the effect of test realism on SI outcomes and second, to evaluate SI of bilateral CI users under realistic conditions. As detailed in the following sections, the levels of realism included (1) standard BKB-like sentences (Bench, Kowal, and Bamford, 1979) in spatially separated standard babble noise, (2) standard BKB-like sentences in three-dimensional recordings of actual situations and (3) the more realistic speech test (RST; Kelly M. Miles et al., 2019) material with the same three-dimensional recordings of actual situations. The evaluation of SI under highly realistic conditions included six different acoustic scenes and were conducted with bilateral CI users tested unilaterally and bilaterally. The effect of test realism was evaluated at 2 different SNRs that corresponded to the SNRs of two of the realistic acoustic scenes. The results of this study will help to better understand the difficulties faced by CI users when communicating in their everyday lives. Moreover, they may provide a more holistic profile of the listening experience of CI users in the real-world and thereby provide a benchmark for evaluating new digital signal processing strategies.

4.2 Methods

As mentioned before, this study has two main goals. First, to evaluate the effect of laboratory-based test realism on SI outcomes of CI users and second, to evaluate SI bilateral CI users tested unilaterally and bilaterally using enhanced ecological validity. This section provides information relevant to the paradigm designed to achieve those two goals, including information about the participants who took part in the study, speech processor considerations, acoustic scenes included, speech material considerations, the criteria to determine their presentation levels, as well as an overview of the main procedures involved in the paradigm.

4.2.1 Subjects and speech processor

Sixteen post-lingually deafened bilateral CI users were initially recruited for the study but only 15 could participate in it, as one of them withdrew due to personal reasons. All participants were users of Cochlear devices, used Nucleus[®]6 speech processors or newer, and were users of the Advanced Combination Encoder (ACE[™]) speech processing strategy (Table 4.1). They all had at least 12 months experience of bilateral electrical hearing, with the exception of subject 12, who only had six months experience.

TABLE 4.1: Biographic data of the participants. ID refers to the identifier of each subject employed throughout this study. Gender is either female (F) or male (M), age is the age of the participant at the time of running the tests, processor is the speech processor model, implant is the type of implant, age at implantation is the age of the participant at the time of the surgery. For an extended version of this table see [Appendix 4.A](#)

ID	Gender	Age	Side	Processor	Implant	Age at implantation
1	F	29	L	N6	CI512	20
			R	N6	CI24RE (CA)	21
2	M	62	L	N7	CI422 (SRA)	56
			R	N7	CI522	59
3	F	52	L	N7	CI24RE (ST)	56
			R	N7	CI24R (CS)	59
4	F	61	L	N7	CI24M	59
			R	N7	CI24RE (CA)	47
5	F	61	L	N7	CI522	59
			R	N7	CI24RE (ST)	47
6	F	45	L	N7	CI512	43
			R	N7	CI24RE (CA)	39
7	F	44	L	N7	CI24RE (ST)	33
			R	N7	CI24R (CS)	26
8	F	61	L	N7	CI24RE (CA)	51
			R	N7	CI512	39
9	M	65	L	N7	CI512_II	63
			R	N7	CI512_II	63
10	M	66	L	N7	CI24R (ST)	45
			R	N7	CI24RE (ST)	32
11	M	60	L	N7	CI512	51
			R	N7	CI512	50
12	F	63	L	N7	CI422 (SRA)	58
			R	N7	CI522	62
13	F	61	L	N7	CI24RE (ST)	47
			R	N7	CI24RE (CA)	51
14	F	48	L	N7	CI422	43
			R	N7	CI422	43
15	M	57	L	N7	CI522	57
			R	N7	CI522	56

During the tests, all subjects wore the same two Nucleus[®]6 speech processors,

provided by CochlearTM. The speech processors were programmed individually by loading the subject's MAP files in Custom SoundTM, which were provided by their audiological clinic. During this process, the following advanced speech processing features were disabled: SCAN, ADRO, WNR and SNR-NR. Although this decision may have moved the speech-in-noise performance obtained in the laboratory away from that seen in the real world, disabling these adaptive features made it easier to evaluate the hearing abilities of the participant, rather than a combination of subject-specific abilities and advanced digital signal processing algorithms. Moreover, turning off the adaptive processing features made it possible to conduct a systematic investigation of the signals presented to the subject during testing. In line with this, the best trade-off between realism, feasibility and control was found by enabling the non-adaptive beamformer available in Cochlear speech processors termed Zoom, which presents a supercardioid pattern (Mauger et al., 2014).

4.2.2 Stimuli

4.2.2.1 Acoustic scenes

The noise signals used in this study consist primarily of three-dimensional recordings of the following noisy situations: Office, Living Room, Church, Dinner party, Cafe and Food court. As further detailed in Sec. 4.2.2.3, these recordings were carried out with a spherical microphone array located in each noisy situation. Additionally, for reference purposes, uncorrelated excerpts from a four-talker babble anechoic signal coming from three different directions (90, 180 and 270 degrees) was included in the study, which was considered here to be a good representative of a relatively complex noise signal that is commonly used in laboratories or clinics where SI is tested with spatially distributed noise sources (Mauger et al., 2014). Table 4.2 provides a summary of all the conditions used in the study. Noise levels varied from 63.3 to 79.4 dBSPL in steps of 3.2 dB and reflected those experienced in the real world. The signal-to-noise ratios (SNRs) were determined by applying the following equation (Eq. 9 in Weisser and Buchholz, 2019):

$$SNR = -16.54 \cdot \log(D) - 0.56 \cdot L + 37.91 + \Delta_G \quad (4.1)$$

where D denotes the talker-to-listener distance in meters, L corresponds to the noise level in dBSPL and Δ_G is a gender correction. In this case, the distance was set to $D = 0.8\text{m}$ and $\Delta_G = -0.84\text{ dB}$, as the target speaker was a female. The (fixed) distance of 0.8m was chosen as a typical communication distance between two interlocutors whose movement is restricted by a small table in between them. Equation 4.1 was obtained after measuring speech levels of natural conversations under different background noise levels and two different interlocutor distances. Applying Eq. (4.1) resulted in SNRs that are assumed to be realistic for the given acoustic environments (see Weisser and Buchholz, 2019). Keeping the distance fixed allowed also to place a loudspeaker in front of the subject at that distance reproducing the

direct sound of the target speech (see Sec. 4.2.2.3). Besides an improved acoustic representation, this provided some useful visual cues for the listening tests (as the loudspeaker was always located at the target position). As summarised in Table 4.2, the resulting SNRs ranged from 3.2 to -5.8 dB, in steps of 1.8 dB. The resulting speech levels (not shown) range from 66.5 dBSPL to 73.6 dBSPL in steps of 1.4 dB.

TABLE 4.2: Summary of all the tested conditions categorised according to their six different SNRs. Six different realistic noises were used as well as multi-talker speech babble, and two different types of speech material: the more realistic speech test (RST) corpus and the "standard" BKB-like corpus. The RST included three different vocal effort levels: soft, moderate and raised. Conditions indicated by an (R) denote that the speech material was convolved with the RIR of each given room.

ID	Speech	Acoustic scene	Noise Level (dBSPL)	SNR (dB)
1	RST soft (R)	Office	63.3	3.2
2	RST soft (R)	Living	66.5	1.4
	BKB-like (R)			
	BKB-like	Babble		
3	RST moderate (R)	Church	69.7	-0.4
4	RST moderate (R)	Dinner	72.9	-2.2
	BKB-like (R)			
	BKB-like	Babble		
5	RST raised (R)	Cafe	76.2	-4
6	RST raised (R)	Food court	79.4	-5.8

The realistic noise environments were all obtained from the same recordings as provided by the ARTE database (Weisser et al., 2019). However, because it was here convenient to sample the presentation levels of the noise (and thus the SNRs) as evenly as possible, the excerpts used here were not exactly the same as the ones available in Buchholz and Weisser (2019). This, along with a subtle final adjustment applied to the presentation levels of the acoustic scenes, resulted in slightly louder environments than those available in Buchholz and Weisser (2019). The different scenes shown in Table 4.2, from soft to loud, correspond to an office (open plan office with keyboard sounds and employees talking), a living room (a scene with a TV sound coming from one of the sides, as opposed to Buchholz and Weisser (2019), and kitchen sounds coming primarily from behind), a small church (a social gathering in a church with several people moving and talking), a dinner party (eight people talking loudly with background music), a cafe (a busy cafe with people talking loudly and loud kitchen sounds) and a food court (a crowded large food court with people talking very loudly). The babble noise was designed as three uncorrelated excerpts of the same four-talker babble coming from three different directions: 90, 180 and

270 degrees (i.e., 12 talkers in total), which is a spatial layout of noise sources that has been previously used in other studies (Mauger et al., 2014).

4.2.2.2 Speech material

Two different types of speech material were used in this study. First, “BKB-like sentences”, which is a corpus developed by the Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation (CRC HEAR) in a similar manner as the Bamford–Kowal–Bench sentences (Bench, Kowal, and Bamford, 1979). The corpus consists of 80 lists of 16 sentences recorded by an Australian female speaker at a sampling frequency of 44.1 kHz. Sentences comprise up to six words or eight syllables and contain vocabulary that is familiar to a five-year-old. Scoring for the BKB-like sentences was done per morpheme, which to our best knowledge is the most common method applied for this speech material.

Second, the newly developed and more realistic sentence test (RST) material (Kelly M. Miles et al., 2019, in preparation). In brief, this speech material was obtained from audiovisual recordings of fluent conversations between two actors. The actors were presented with three different noise levels via open headphones while talking to each other. In this way, conversations at three different vocal efforts levels were recorded, which are referred to here as “RST soft”, “RST moderate” and “RST raised”. The level of the noise presented via the open headphones was 60 dB SPL (soft), 71 dB SPL (moderate) and 80 dB SPL (raised). The conversations were cut into short sentences with a length from three to 12 words and an average of 6 words. The sentences were intelligibility normalised by measuring individual psychometric functions for each sentence in young normal-hearing listeners, and then applying a sentence specific gain that compensated for the differences in the shift of these psychometric functions. As part of the same process, sentences were grouped into 12 equivalent lists of 16 sentences each, with 4 lists per vocal effort level.

In this study, the speech material of each of the vocal effort levels was used in two acoustic scenes (Table 4.2) but slightly different sound pressure levels. “RST soft” was used in the Office and in the Living Room, “RST moderate” was used in the Church and in the Dinner party, and “RST raised” was used in the Cafe and in the Food Court. In each acoustic scene, one list of 16 sentences was used for unilateral and another one for bilateral testing. In contrast to the BKB-like sentences, the RST sentences were scored per word and not per morpheme. This was necessary because due to the fluent spontaneous speech the number of morphemes per sentence was too large and per morpheme scoring therefore not feasible.

4.2.2.3 Sound reproduction

The stimuli were presented to the listeners via a spherical array of 41 loudspeakers (Tannoy V8) placed on a sphere with a radius of 1.85 m and located inside the anechoic chamber of the Australian Hearing Hub, Macquarie University (Weisser et al.,

2019). The noise signals described in Sec. 4.2.2.1 were recorded with a 62-channel microphone array and decoded into 41 loudspeaker signals using the higher-order Ambisonics (HOA) method (see Weisser et al., 2019 for details). This resulted in a highly authentic reproduction of the original acoustic scenes at their original sound pressure levels.

In order to provide target speech with realistic scene-specific reverberation, HOA-encoded Room Impulse Responses (RIR) for each of the acoustic environments were obtained from the ARTE database and decoded into the 41 channels of the playback loudspeaker array. As the intended talker-to-listener distance was 0.8m (see Sec. 4.2.2.1), the direct sound component of the multi-channel RIRs was extracted and presented via an additional loudspeaker (Genelec 8020C) that was placed in front of the subject at a distance of 0.8 m. Since the RIRs were recorded at a distance of 1.3 m, the amplitude of the direct sound component was adjusted by a factor of $1.3/0.8$. The rest of the RIR was kept untouched and presented via the 41-channel loudspeaker array. The justification for such an approximation stems from the assumption that the reverberation field is distance independent and the pressure of the direct sound component follows the one-over-distance rule for omni-directional sound sources.

The frequency response of all Tannoy V8 loudspeakers was equalized using minimum-phase FIR filters, and differences in sensitivity and acoustic delay with the nearby Genelec 8020C loudspeaker was compensated.

4.2.3 Procedures

Subjects were tested both unilaterally and bilaterally. Due to the limited number of sentence lists provided by the RST speech material, most unilateral conditions were only tested with the better-performing ear (BPE). This decision was made to avoid attributing any potential bilateral benefit to the addition of a better performing ear. For control purposes, two conditions were additionally tested with the worse performing ear (WPE) using the BKB-like sentences (Table 4.2). The BPE was determined by evaluating the unilateral Speech Reception Threshold (SRT, Keidser et al., 2013) separately for each ear at the very beginning of the experimental session. The SRTs were measured with BKB-like sentences in the presence of collocated 12-talker babble noise. The babble-noise signal was the same as the one described in Sec. 4.2.2.1 except that in this case all three 4-talker babble noise signals came from the same loudspeaker as the target speech. Two SRTs per ear were measured and the two values averaged. In case the difference between the scores of the left and right ears was lower than 3 dB, the subject's preferred ear was chosen as the BPE. Otherwise, the ear with the lower SRT was chosen. The results, along with the tested ear (i.e., the ear assumed to be the BPE) are shown in the appendix (Table 4.B.1).

Because the acoustic scene was not left-right symmetrical, and the BPE varied across subjects, two different versions of each acoustic scenes were created, one being the original and the other one being a left-right flipped (mirrored) version of it. In this way, it was ensured that all subjects were presented with an identical scene relative

to their BPE. However, the test paradigm did not control for the ear side with the better SNR. Hence, as shown in the next sections, the BPE corresponded to the ear with better SNR only in a subset of conditions.

During the test, once the SRTs were measured and the BPE was determined, a practice trial comprising 16 sentences in the RST-VSE condition was administered before the main experiment was started to enable participants to familiarise themselves with the new speech and noise materials. Thereafter, the main experiment was started by testing subjects with their BPE (Table 4.2). As the whole test lasted about two hours, all subjects were tested unilaterally first to avoid attributing potential fatigue effects to the bilateral benefit. After finishing all the ten conditions to be tested with the BPE (see Table 4.2), subjects were asked to take a break of approximately ten minutes. The test was then resumed with the two WPE conditions. Right after, subjects were asked to turn on their second device and the ten remaining conditions were tested bilaterally. Within each of the blocks (i.e. BPE, WPE and bilateral) conditions were randomized.

4.2.4 Instrumental signal evaluation using the U50

In order to have a better understanding of the degree to which the differences in SI between conditions can be explained by background noise and reverberation at the pre-processing stage of the speech processor, the U50 (Bradley, 1986) was evaluated at the output of the speech processor's (non-adaptive) beamformer. The U50 is a SNR measure that is commonly used for reverberant speech, in which the early reflections and the direct sound components of the target speech are assumed to contribute positively to speech understanding, whereas the late reflections add to the power of the noise and limit the understanding of speech (see Chapter 4, page 104 in Kates, 2008). The U50 is defined as:

$$U50 = 10 \cdot \log \left(\frac{DSER}{LR + N} \right), \quad (4.2)$$

with DSER the power of the direct sound and early reflections of the reverberant target speech combined, LR the power of the late reflections, and N the power of the noise.

The first step to calculate the U50 in each condition consisted of simulating the speech and noise signals, separately, at the speech processor's front and rear microphones. This was done by measuring a set of impulse responses (IRs) from each of the 41 loudspeakers in the playback array to each of the two microphones of a speech processor placed on the left and right ear of a HATS located in the center of the loudspeaker array. Thereby, a modified version of a commercially available speech processor was used that enabled access to the signal at the output of the microphones. Once the four impulse responses were obtained for each loudspeaker, calibrated simulations of the acoustic scenes and (reverberant) target speech signals

were derived for each microphone individually by convolving these IRs with the target speech material and noise files used within the actual speech tests. For the target signals, separate simulations were conducted for the early (plus direct sound component) and the late reflections. Before applying the beamformer filter that combined the front and rear microphone signals of each speech processor to generate the directional response of a supercardioid microphone (see Sec. 4.2.1), the signals were downsampled to match the sampling frequency of a commercial speech processor (around 16 kHz). After applying the beamformer filter, the signal corresponding to the early reflections plus direct sound component was readily available for the calculation of the U50, and the signal corresponding to the effective noise was obtained by adding the noise signal and the late reflections. As opposed to the presentation SNR given in Table 4.2, which were calculated based on the broadband RMS of the signal, the calculation of the U50 was based on their dB SPL value in third octave bands from 125 Hz to 8 kHz. The final expression of the U50 consisted of the mean across frequency bands of the difference in dB between the useful and the detrimental parts.

4.2.5 Statistical analysis

In the present study, several statistical models were applied to the data, which varied from case to case as further described in the results section. In this section, only the aspects that are common to the individual analyses are described. First, the SI data was linearized by means of the following transformation (Warton and Hui, 2011):

$$SI_{lin} = \ln\left(\frac{SI_n + \epsilon}{1 - SI_n + \epsilon}\right) \quad (4.3)$$

with SI_{lin} the linearised SI data, SI_n the raw SI data normalised to one and ϵ the minimum non-zero value of $(1 - SI_n)$ across all SI scores, as suggested in Warton and Hui (2011). Second, a linear mixed effects model was fitted and the residuals were checked for normality with a Shapiro-Wilk test. Third, outliers were identified and removed. Data points were considered outliers if their standardised residual was located beyond 2.5 standard deviations from 0. Fourth, if required, the statistical model was fitted again to the data excluding outliers. Finally, an analysis of variance (ANOVA) was conducted from the output of the model. The linear mixed-effects models were realised with the R package *lme4* (version 1.1-17), and their fitting was conducted following the Bound Optimisation BY Quadratic Approximation (bobyqa) algorithm. Outlier removal was conducted with the *romr.fnc* function available in *LMERConvenienceFunctions* package (version 2.10). The ANOVA analyses were conducted with the package *lmerTest* (version 2.0-36). T-tests were conducted with the *emmeans* package and multiple comparisons were Tukey-corrected.

4.3 Results

Figure 4.1 shows a summary of all the SI data obtained unilaterally with the BPE (top panel) and bilaterally (bottom panel). The data shown includes SI scores obtained in the six realistic RST-VSE conditions at six different SNRs, as well as in the BKB-VSE and BKB-Babble conditions at two different SNRs. Speech intelligibility is consistently higher in BKB-babble than in BKB-VSE, which in turn, is consistently higher than in RST-VSE. Speech intelligibility in RST-VSE gradually increases with increasing SNR from -5.8 dB to -0.4 dB, where it plateaus. Speech intelligibility scores are slightly higher in the bilateral than in the unilateral conditions, although from the results shown in Fig. 4.1 it is not clear how consistent this bilateral benefit is across subjects.

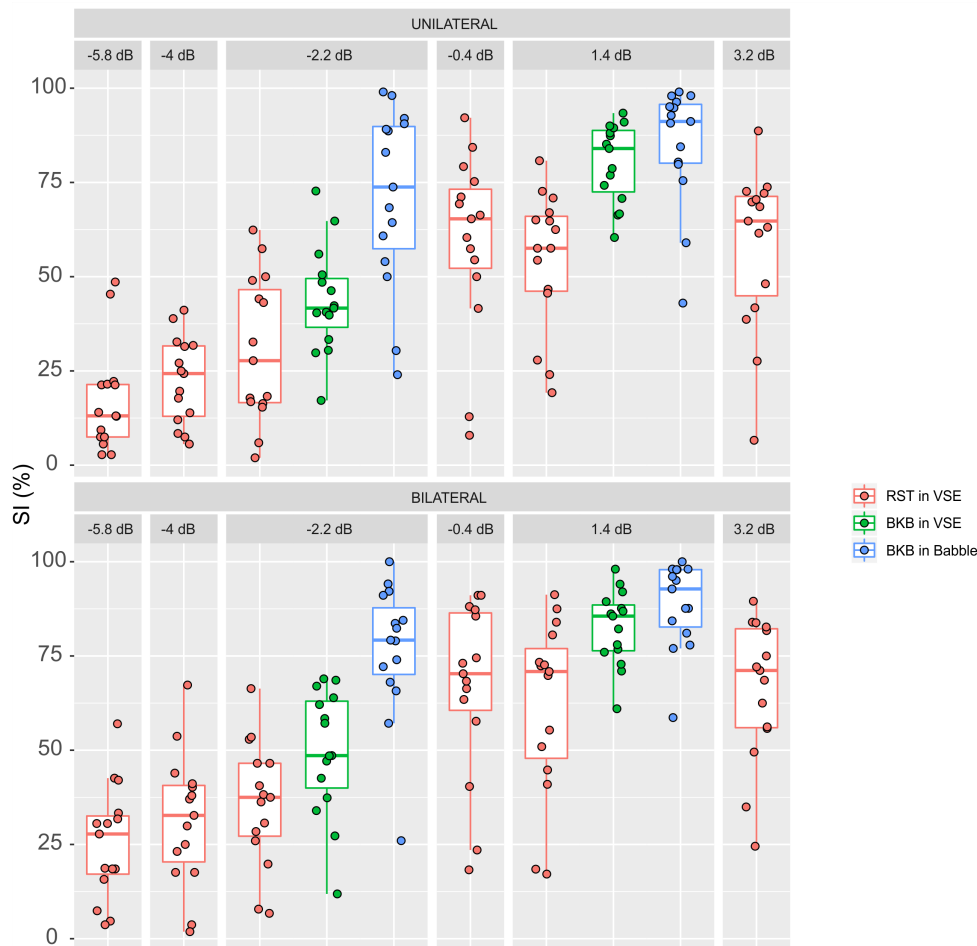


FIGURE 4.1: Median SI scores, interquartile range and raw data obtained at the BPE (top) and bilaterally (bottom) at six different SNRs considering RST sentences in the six VSEs. Speech intelligibility scores for BKB sentences in the dinner party (SNR = -2.2 dB), the living room (SNR = 1.4 dB) and in babble noise are also shown (see text for details).

The following sections are divided into three main parts. The first part is concerned with the effect of realism on speech intelligibility outcomes and the observed

bilateral benefit. The second part presents the SI outcomes measured in the different realistic conditions both unilaterally with the BPE and bilaterally. The third part of this section conducts a further evaluation of the bilateral benefit seen in the previous parts by additionally taking into account the WPE.

4.3.1 Effect of test realism on SI outcomes

As described before, one of the goals of the present study is to evaluate the effect of the level of realism of both noise and speech stimuli on SI outcomes. Thereby, three different levels of realism were considered: "standard" BKB-like sentences in "standard" babble noise (BKB-Babble), "standard" BKB-like sentences in realistic VSEs (BKB-VSE) and more realistic RST sentences in the realistic VSEs (RST-VSE). Fifteen participants were tested at these 3 levels of test realism at two different SNRs (1.4 dB and -2.2 dB) in both unilateral (BPE only) and bilateral listening mode. Figure 4.2 shows the mean SI data and 95% confidence intervals for the unilateral (solid circles) as well as for the bilateral mode (solid triangles) as a function of the level of test realism at an SNR of 1.4 dB (upper panel) and -2.2 dB (bottom panel). As can be seen, overall performance drops for increasing levels of realism (i.e., from left to right) and decreasing SNR for both the unilateral (BPE) and the bilateral listening mode.

A linear mixed-effects model (with random intercept for subject with random slope for level of realism) was applied to the linearised SI scores to infer any statistically significant effect of SNR, level of realism, listening mode (unilateral or bilateral), as well as the interaction between SNR and level of realism (see Sec. 4.2.5 for more details). F tests conducted from the model revealed a significant effect of level of realism [$F(2,15.4) = 61.1$; $p < 0.001$], SNR [$F(1,142.4) = 328.9$; $p < 0.001$], their interaction [$F(2,142.4) = 11.1$; $p < 0.001$], and listening mode [$F(1,142.4) = 17.7$; $p < 0.001$]. Hence, although the bilateral SI scores were only slightly higher than the unilateral scores, the bilateral benefit was significant overall.

The effect of the realism of the noise material can be derived by comparing the SI scores in BKB-Babble and BKB-VSE shown in Fig. 4.2. This difference in SI is higher for lower SNRs. However, the average SI scores of BKB-babble at 1.4 dB SNR are located in a less sensitive (or shallower) region of the performance intensity function than those obtained at -2.2 dB SNR. As a consequence, the actual value of these differences is a by-product of the effect of SNR on SI along with the location on the performance intensity function at that SNR. Hence, the effect of noise material is only part of the reason why the drop in performance observed from the rather artificial babble noise to the more realistic VSEs is higher at -2.2 dB SNR than at 1.4 dB SNR. Similar observations can be made for the effect of the noise on the bilateral data (Fig. 4.2, solid triangles), except that the bilateral SI scores are consistently higher than for the BPE.

The effect of the realism of the speech material (along with the potential effects of different female speakers in each case) can be derived by comparing the SI scores in BKB-VSE and RST-VSE shown in Fig. 4.2. In this case, the drop in performance

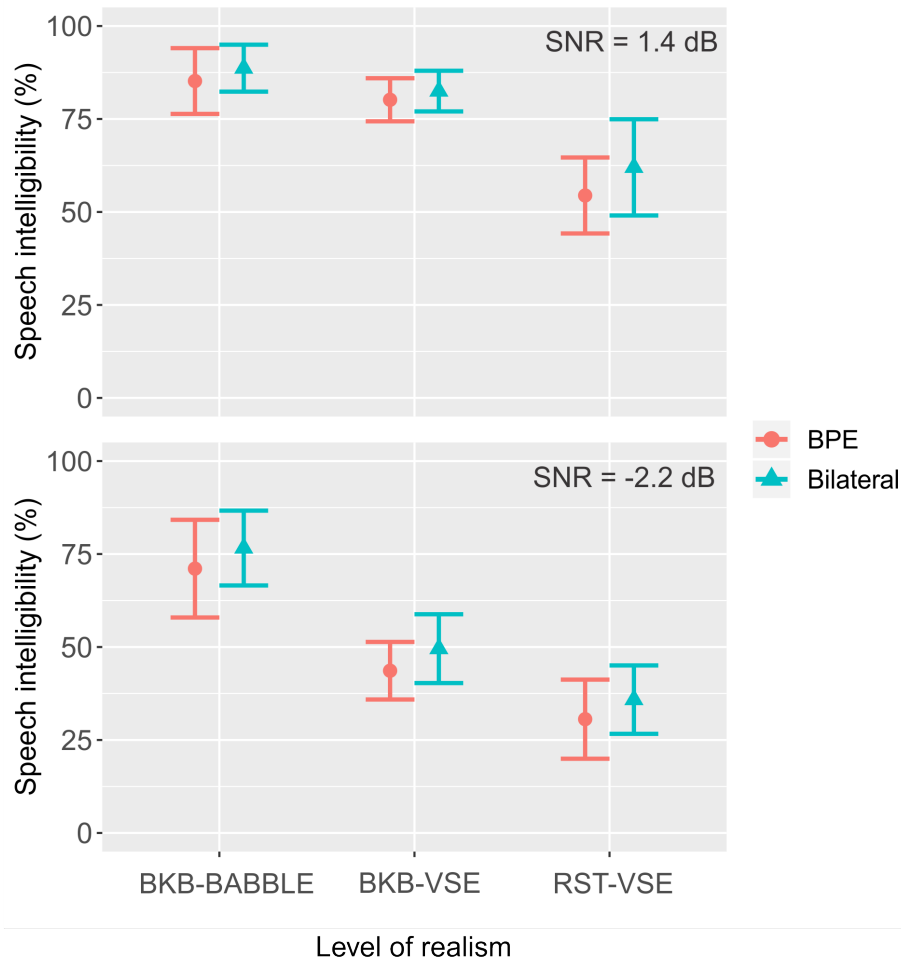


FIGURE 4.2: Mean speech intelligibility scores with 95% confidence intervals measured at 1.4 dB SNR (upper panel) and at -2.2 dB SNR (bottom panel) in three different levels of realism: BKB-Babble, BKB-VSE and RST-VSE (see text for more details).

from BKB sentences to the more realistic RST sentences is clearly higher at an SNR of 1.4 dB than at -2.2 dB, even though the average data at 1.4 dB is in a slightly shallower (or less sensitive) region of the performance intensity function which further pronounces this difference. Speech intelligibility at 1.4 dB SNR drops by about 25% (from 80% to 55%) whereas at -2.2 dB SNR, it drops by less than 15% (from 43% to 31%). This indicates that at higher (positive) SNRs, where the target speech dominates, SI depends more strongly on the sentence material than at lower (negative) SNRs, where the noise dominates. Similar observations can be made for the effect of the speech material on the bilateral data (Fig. 4.2, solid triangles), except that the bilateral SI scores are consistently higher than for the BPE.

Comparing the relative effects of the realism of the speech material and background noise on the SI scores shown in Fig. 4.2, it can be seen that the effect of the noise is more pronounced at the lower (negative) SNR of -2.2 dB than in the higher (positive) SNR of 1.4 dB, whereas the opposite can be observed for the effect of the speech material, which is more pronounced at the higher (positive) SNR. However,

these effects are also influenced by the absolute position of the SI scores on the performance intensity function (see above), as well as by a potential interaction of the inherent changes in speech material and noise condition, i.e., the effect of the noise material is evaluated with the BKB sentences, but not the RST sentences, whereas the effect of the sentence material is evaluated in the VSEs but not in the babble noise.

The effect of the three levels of realism on SI and their dependence on the SNR is further analyzed in Fig. 4.3 with a particular focus on the behaviour of the individual data. For simplicity, only the unilateral (BPE) data is shown here, but very similar observations can be made for the bilateral data. The upper-left panel shows the three possible combinations of the speech (RST or BKB) and noise material (Babble or VSE) that is included in the analysis and make up the three levels of test realism. The other three panels (panels A, B, and C) show the SI scores re-plotted from Fig. 4.1.

Panel A shows the effect of the noise using BKB-like sentences. In this case, the only difference between the axes is the noise material and, as already described above, the SI scores in the VSE are consistently worse than in the babble noise. Moreover the SNR has a small effect in the babble noise but a large effect in the VSE (where the change in SNR is accompanied by a change in noise signal), which can be observed from the slopes of the straight lines that connect the individual and mean SI scores at the two SNRs, which are all much greater than one.

Panel B shows the effect of the speech material in the more realistic VSEs. In this case, the only difference between the two axes is the speech material. The SI scores are consistently lower for the RST than for the BKB material. As opposed to panel A, the slope of the solid black line is only slightly lower than one, suggesting that the effect of the SNR is slightly lower in RST-VSE than in BKB-VSE. As shown in the last panel of Fig. 4.2, this is primarily due to the fact that at the high SNR of 1.4 dB, the mean SI scores for the BKB-VSE condition were close to 80% whereas those obtained in RST-VSE were close to 55%, and were thus at a shallower region of the performance intensity function. This is in contrast to the lower SNR of -2.2 dB SNR, for which SI scores of the BKB and RST material tend to converge, as they are both highly dependent on the noise level. This explains why the effect of the SNR is slightly lower in RST-VSE than in BKB-VSE (see also upper panel of Fig. 4.4, where the effect of SNR is shown in each case).

Panel C shows the combined effects of speech and noise materials. The SI scores for RST-VSE are plotted against those for the BKB-Babble. Speech intelligibility scores are consistently lower in the RST-VSE than in the BKB-Babble condition. In this case, the effect of the SNR on SI is larger in the RST-VSE (where it is accompanied by a change in noise signal) than in the BKB-Babble, as the slopes of the straight lines are generally higher than one.

Since the SNRs considered in the above analyses were calculated from the broadband levels of the speech and noise stimuli as picked up by an omni-directional microphone, it is unclear how far the differences observed between the three levels of test realism can be explained already by the more detailed (in particular spectral)

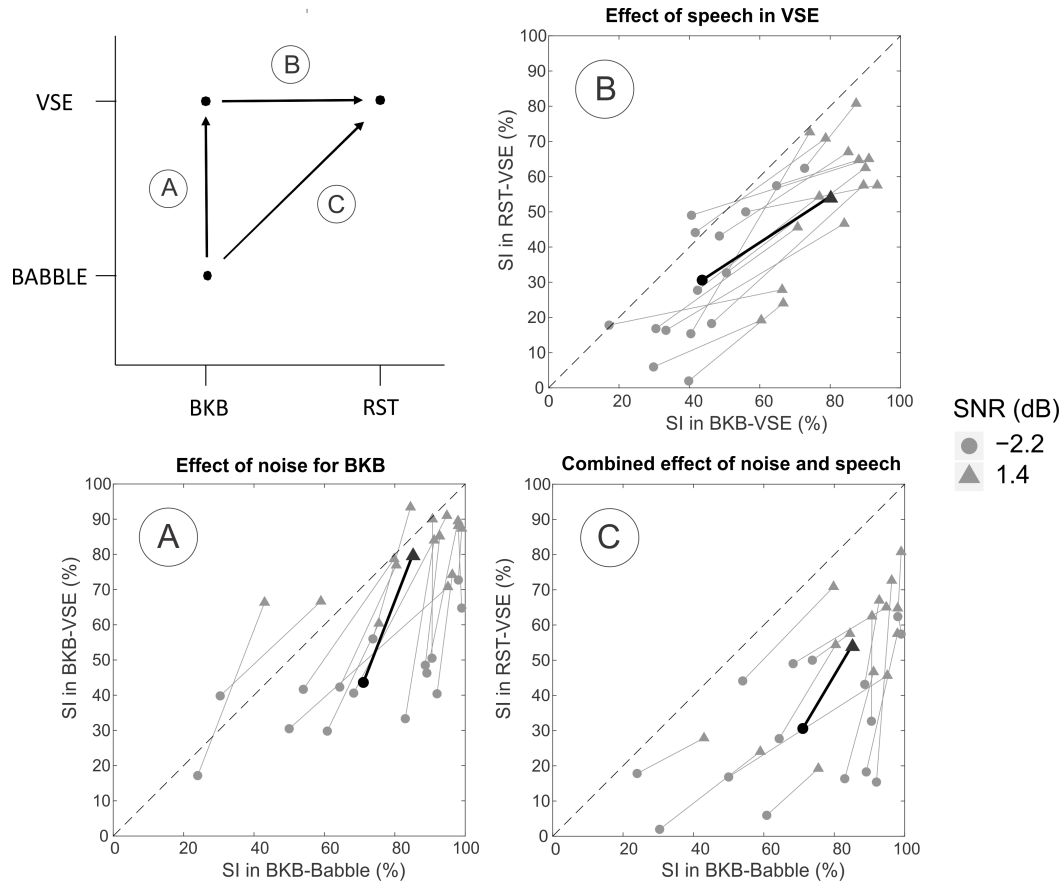


FIGURE 4.3: Effect of test realism on SI outcomes. The first panel illustrates the connection between the three stimuli conditions to be compared. The other three panels show SI data measured with each subject in two different stimuli conditions, as indicated by their axes. Each of the panels shows the SI obtained by each subject (lines) at two different SNRs (markers). Panel A compares SI in BKB-Babble against that of BKB-VSE. Panel B compares SI data of BKB-VSE against that of RST-VSE. Panel C compares SI data of BKB-Babble against SI of RST-VSE.

characteristics of the signals that arrive at the directional input of the subjects' speech processors during testing. Therefore, the SI data was further analyzed by evaluating the U50 for the target and noise stimuli used in the different test conditions at the output of the (non-adaptive) beamformer of the subject's speech processors, as explained in Sec. 4.2.4. Figure 4.4 shows the mean SI scores with 95% confidence intervals obtained in each of the three levels of realism for the two tested SNRs (connected lines) as a function of the broadband SNR (upper panel) as well as of the U50 at the directional input of the subjects' speech processors (bottom panel). Figure 4.4 suggests that the SI scores for the BKB-VSE and BKB-Babble are part of the same performance intensity function when plotted as a function of the U50, which is not the case when plotted as a function of SNR. In contrast, RST-VSE presents much lower SI scores than any of the other levels of realism in both figure panels. This indicates that the U50 can largely explain the differences in SI observed between noise types, but it is not able to explain the differences in SI observed between speech materials.

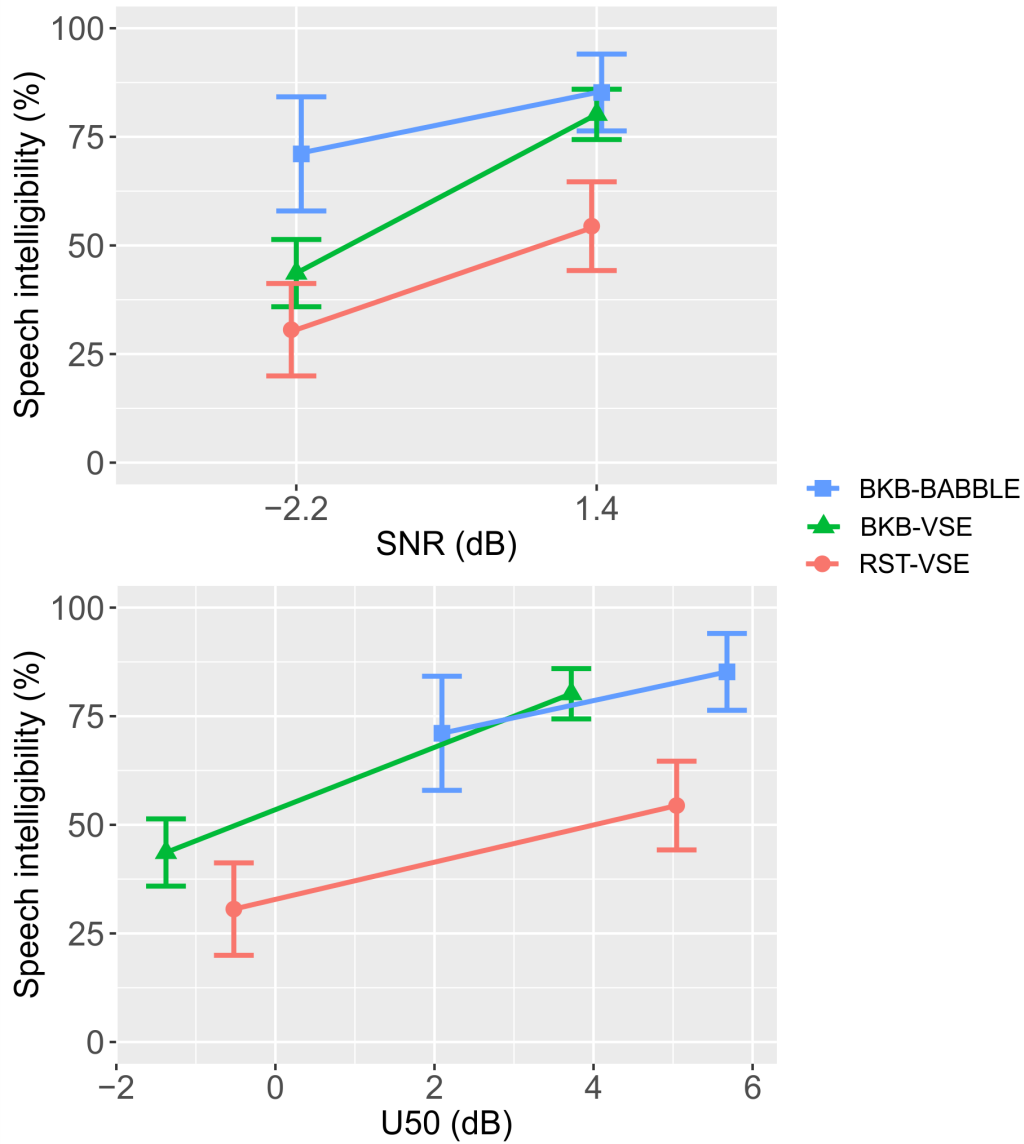


FIGURE 4.4: Mean SI scores with 95% confidence intervals obtained in each of the speech-noise configurations (markers and connecting lines) as a function of the U50 at the output of the beamformer.

To understand how far the U50 could explain the differences seen in the data, a linear mixed-effects model was applied to the linearised data with level of test realism and U50 as predictor variables and a random intercept for subjects. F tests revealed a significant effect of the level of realism [$F(2,69.07) = 20.68$; $p < 0.001$] and U50 [$F(1,69.05) = 94.6$; $p < 0.001$], but no significant interaction [$F(2,69.03) = 2.36$; $p = 0.1$]. A t-test with Tukey correction showed that RST-VSE was significantly different from both BKB-VSE [$t(69) = -8.5$; $p < 0.001$] and from BKB-Babble [$t(69.1) = -8.5$; $p < 0.001$]. As expected from the results shown in Fig. 4.4, BKB-VSE and BKB-Babble were not significantly different from each other [$t(69.1) = -0.8$; $p = 0.69$].

4.3.2 Speech intelligibility in realistic conditions

As shown in Fig. 4.1 (red symbols), SI scores in the six RST-VSE conditions with the realistic speech material as well as noises (see Table 4.2) were on average quite low. For the lowest SNR of -5.8 dB (the food court scene), the median SI score for the unilateral (BPE) condition was 13%. The median SI performance increased slowly with increasing SNR up to -0.4 dB SNR (the dinner party scene) and then plateaued at an average performance of about 63%. Even though all subjects showed the same trend, the individual performance varied substantially. Whereas the best performers reached SI scores of more than 90% at the highest SNR conditions, some never exceeded SI scores of more than 20%. The bilateral data presented the same trends, except that the overall SI scores were slightly higher.

A linear mixed-effects model was applied to the RST-VSE data with SNR and listening mode (unilateral or bilateral) as fixed effects and subjects as random intercepts with random slopes for listening mode (to account for the subject-level effect of listening mode). F tests revealed a significant effect of SNR [$F(5,150)=140.58$; $p<0.001$], a significant effect of listening mode [$F(1,150)=35.96$; $p<0.001$] and no significant interaction [$F(5,150)=0.52$; $p = 0.76$]. A series of t-tests (Tukey-corrected for multiple comparisons) revealed that SI at the highest three SNRs were not significantly different from one another, neither for the unilateral nor the bilateral listening mode. In particular, t-tests, averaged over listening mode, showed [$t(140) = 1.9$; $p = 0.4$] for 3.2 dB vs 1.4 dB, [$t(141) = -0.7$; $p = 0.98$] for 3.2 vs -0.4 dB and [$t(140) = -2.6$; $p = 0.09$] for 1.4 vs -0.4 dB.

In order to evaluate the potential bilateral benefit in each acoustic scene separately, a series of t-tests were conducted conditioned on SNR. The results showed a significant effect of listening mode in the office [$t(113) = -2.1$; $p < 0.05$], the living room [$t(111) = -2.4$; $p < 0.05$], the café [$t(116) = -2.4$; $p < 0.05$] and in the food court [$t(111) = -3.5$; $p < 0.05$]. In contrast, no significant effect of listening mode was found in the church [$t(113) = -1.9$; $p = 0.05$] and in the dinner party [$t(113) = -1.4$; $p = 0.16$].

Even though the statistical model revealed a significant difference between the unilateral and bilateral SI scores, this difference (i.e., the bilateral benefit) is hard to see in Fig. 4.1 due to the large variance across subjects. To further evaluate the bilateral benefit received by the individual subject, Fig. 4.5 (top panel) shows the subjects' individual unilateral (BPE) SI scores versus their bilateral scores for all six SNRs included in the RST-VSE condition. The regression line in Fig. 4.5 estimates a 7% SI improvement on average of bilateral over unilateral scores throughout the linear region of the underlying sigmoidal performance intensity function (i.e. around 50%). To better understand the effect of the SNR (or acoustic environment) on the bilateral benefit, the middle panel in Fig. 4.5 shows the SI scores averaged across subjects in each of the six SNRs. The average bilateral benefit is again around 7% and is rather constant across SNRs. To better understand the subject-specific bilateral benefit, the bottom panel in Fig. 4.5 shows the SI scores averaged across SNRs separately for

each subject. In this case there is more variability than in the previous panel, which may indicate that the bilateral benefit presents a higher variability across subjects than across SNRs. This can be confirmed by calculating the standard deviation of the bilateral benefit, which is 1.3% for the SNR effect and 4.9% for the subject effect. The average bilateral advantage is in this case predicted to be slightly higher than in the previous two cases, with an estimate of 8.3%, which is most likely due to the performance intensity function underlying the SI scores that can only be assumed to be linear around the 50% point. The regression lines presented in the three panels of Fig. 4.5 have slopes very close to unity. This suggests that the bilateral benefit does not depend on the SI performance attained with only one ear.

In order to better understand the results obtained in the RST-VSE conditions (both unilaterally and bilaterally), in Fig. 4.6 the mean SI scores with 95% confidence intervals for the six RST-VSE conditions are plotted as a function of the U50 at the output of the speech processor's beamformer (instead of the broadband SNR measured in free-field). For the bilateral case, the U50 corresponds to the side with the more favourable U50.

Figure 4.6 partly explains the reason why the three acoustic scenes with the highest SNRs (i.e., the office, the living room and the church) had similar SI scores. As opposed to their broadband SNRs measured in the free-field, which spans a range of 3.6 dB, their U50s at the output of the beamformer spanned a significantly narrower range of about 2 dB.

With respect to the measured bilateral benefit, Fig. 4.6 confirms that, irrespective of whether the unilateral test was conducted at the ear with most favourable U50 or not, there was only a small bilateral benefit in terms of SI. However, the dinner party (data points around 0 dB) stands out for having a more favourable U50 value at the WPE than at the BPE.

A linear mixed-effects model with listening mode and U50 as predictor variables and subjects as a random intercept followed by an F-test, showed a significant effect of both listening mode [$F(1,41.3) = 16.3$; $p < 0.001$] and U50 [$F(1,159) = 537.6$; $p < 0.001$], but no significant interaction [$F(1,159) = 0.19$; $p = 0.66$].

4.3.3 Further analysis of bilateral SI advantage

Due to the limited number of RST sentence lists that were available (Sec. 4.2.2.2), only the BPE could be tested in the six RST-VSE conditions in addition to the bilateral listening mode. As a consequence, the bilateral benefit observed in Sec. 4.3.2 was calculated by the bilateral minus the unilateral (BPE) SI scores. Therefore, it is unclear if in some conditions the assumed BPE was actually the poorer performing ear, and the assumed WPE could in fact have provided better SI outcomes. Therefore, the bilateral benefit seen in at least some of the RST-VSE conditions given in Sec. 4.3.2 may overestimate the true bilateral benefit. To address this potential issue, the WPE was measured in addition to the BPE as well as bilaterally in the living room (1.4 dB SNR) and in the dinner party (-2.2 dB SNR) environments using BKB-like sentences. As a

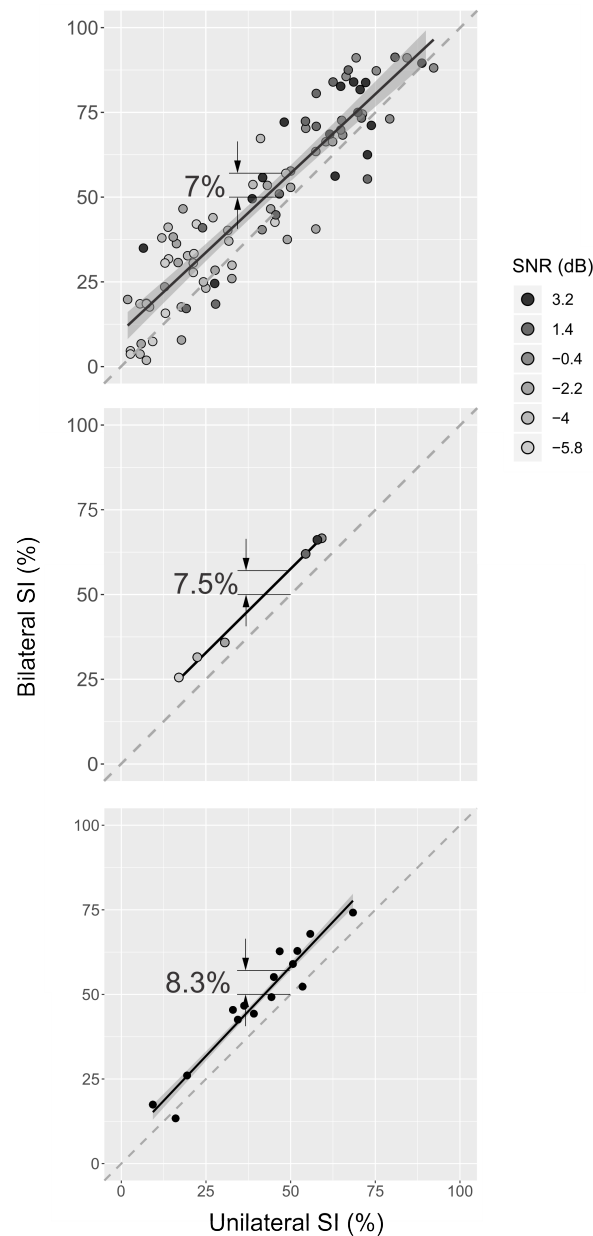


FIGURE 4.5: Scatter plots and regression lines with 95% confidence intervals of SI scores obtained bilaterally as a function of SI scores obtained unilaterally. The first panel shows the raw data obtained by all subjects in all SNRs. Second panel shows the mean scores across subjects per condition. The third panel shows the mean scores across conditions per subject. The percentages shown in each panel correspond to a rough estimate of bilateral SI advantage.

reminder, when tested with the RST speech material, these acoustic scenes led to a significant and a non-significant effect of listening mode in the living room and in the dinner party respectively. The resulting data was analysed by a linear mixed-effects model with the listening mode (WPE, BPE, and bilateral) and SNRs (or acoustic environments) as fixed effects and a random subject specific intercept with random slope for listening mode. F tests in this case showed that there was a significant effect of both SNR [$F(1,56) = 413.1$; $p < 0.001$] and listening mode [$F(2,20.1) = 15.9$; $p < 0.001$]

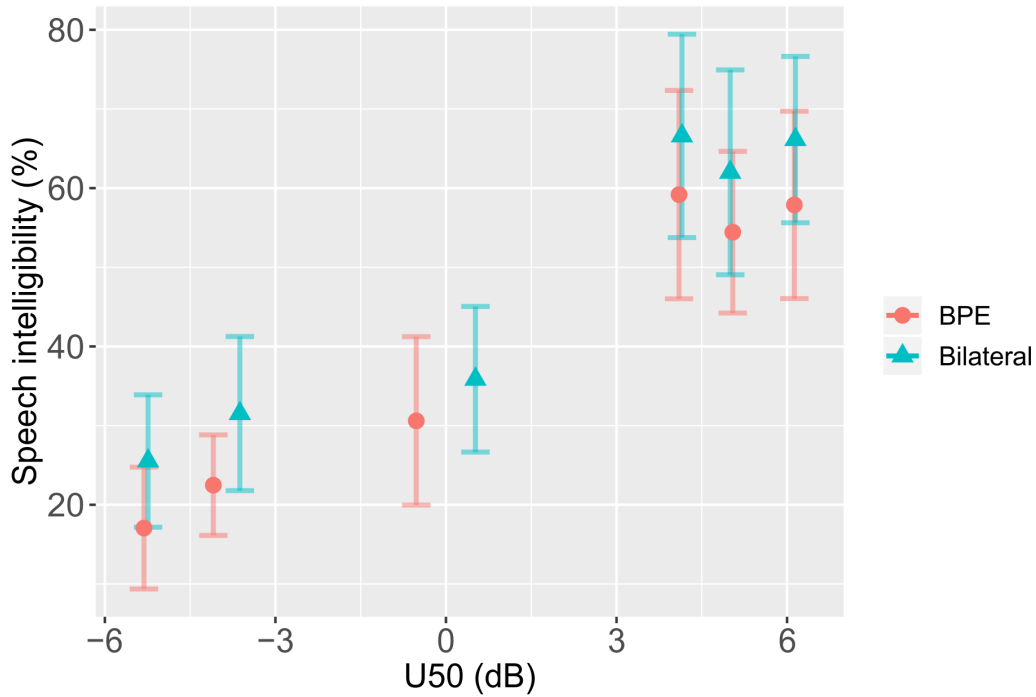


FIGURE 4.6: Mean of SI scores across subjects with 95% confidence intervals of the RST-VSE conditions for the BPE (solid circles) and bilateral (triangles) listening modes. The results are plotted as a function of the U50. In the bilateral case, the U50 of the better-U50 ear was taken.

but no significant interaction [$F(2,56) = 1.01$; $p = 0.37$]. However, a series of t-tests showed that in none of the SNRs there was a significant difference between the BPE and the bilateral mode ($t(34.3) = -1.24$; $p = 0.44$ for 1.4 dB SNR and $t(34.3) = -1.44$; $p = 0.33$ for -2.2 dB SNR). Hence, in regards to the bilateral benefit obtained with respect to the BPE, the results obtained with the BKB-like sentences agree with those obtained with the RST sentences in the dinner party and disagree in the living room.

In the BKB-VSE data, a subtle difference was observed between the two SNRs (acoustic scenes). At -2.2 dB SNR (i.e., in the dinner party), a t-test between the WPE and the BPE showed no significant difference ($t(19.6) = -2.46$; $p = 0.06$). However, at 1.4 dB SNR (i.e., in the living room), the BPE and WPE were significantly different from each other ($t(19.6) = -3.46$; $p < 0.05$).

In order to understand the results of the previous statistical analysis, Fig. 4.7 shows subject-averaged SI scores with 95% confidence intervals of the six conditions as a function of the U50 at the output of the beamformer of the subjects' corresponding speech processors, or at the ear with most favourable U50 for the bilateral condition. It can be observed that the reason why the WPE and BPE are significantly different in the living room (i.e., at 1.4 dB SNR) is primarily due to the fact that the WPE has a much lower U50 than the BPE. This was due to commercials that were presented from a TV at the side of the WPE and thereby provided the WPE with a less favourable U50. In contrast, although Fig. 4.7 suggests that the WPE can deal with noise substantially worse than the BPE in the dinner party environment (-2.2

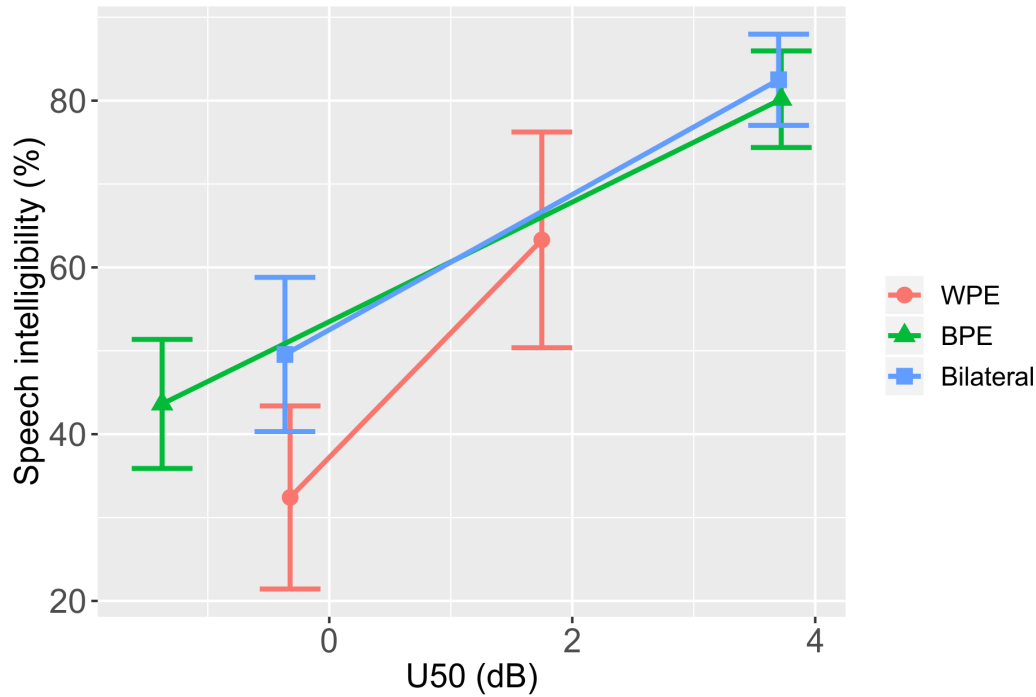


FIGURE 4.7: Mean of SI scores across subjects with 95% confidence intervals of the BKB-VSE conditions for the WPE, the BPE and bilateral listening mode. The results are plotted as a function of the U50. In the bilateral case, the U50 of the better-U50 ear was taken.

dB SNR), the WPE has a more favourable U50 than the BPE, which explains why the scores between these two conditions are not significantly different. Figure 4.7 also suggests that the SI performance with two ears is not significantly different from that obtained with the BPE, which was already confirmed by means of t-tests. However, it is not clear whether this is a consequence of the near-ceiling effects in the living room or the fact that, in the dinner party, the BPE corresponded to the side with the worse U50.

4.4 Discussion

The main goals of this study are to understand the effect of the level of test realism on SI outcomes in bilateral CI users as well as to measure absolute SI performance in more realistic laboratory conditions. The discussion section is organised around these two goals.

4.4.1 Effect of test realism on SI

The levels of realism applied in the present study included BKB-Babble, BKB-VSE and RST-VSE, which differed in their level of realism for the applied noise (Babble versus VSEs) and speech material (BKB-like versus RST sentences). Speech intelligibility was shown to decrease significantly with increasing level of realism for both of the considered SNRs (i.e., -2.2 dB and 1.4 dB), which for the more realistic noise

was realised by two different virtual sound environments (VSEs). Interestingly, at the higher SNR of 1.4 dB, where the speech signal had a higher level than the noise, the BKB-like sentences led to far higher SI scores than the RST sentences. In contrast, at the lower SNR of -2.2 dB, where the noise had a higher level than the target speech, this difference almost disappeared and both speech materials led to similar SI. The opposite occurred with the effect of the noise material. In this case, SI scores measured with the rather artificial babble noise and the more realistic VSEs were rather similar at higher (positive) SNR, whereas at the lower (negative) SNR, scores measured in the VSEs were much lower. Hence, the data analysis revealed a significant interaction between the effect of the level of test realism and the SNR.

However, an instrumental analysis of the sound stimuli that were applied during testing showed that the differences in SI between the babble noise and the VSEs that was seen at both SNRs could be explained by the U50 measured at the output of the (fixed) beamformer of the subjects' speech processors. The U50 showed that babble noise led to much more favourable U50 values than the VSEs even though their broadband SNRs measured in free-field were the same. Since the target speech material was BKB-like sentences in both conditions and only differed by a small amount of reverberation, this was likely the effect of the directional characteristics of the beamformer, which was much more effective in reducing the level of the noise in the babble noise than in the two VSEs. However, the U50 was not able to explain the substantial differences in SI that were observed between the BKB and RST sentences inside the VSEs. This may be expected, as the U50 only accounts for spectral (long-term) differences and the overall level of the target speech, and omits other important factors such as syllabic rate, context or articulation.

In order to provide information about the extent to which the SI outcomes measured in a given stimulus configuration may be generalised to other configurations, the coefficient of determination (R^2) was calculated across all the possible combinations. The results are shown in Fig. 4.8 as a correlation diagram. In general, the correlation coefficients between cases are moderate, indicating that subjects who performed better than the rest in a given condition do not necessarily perform consistently better in a different condition. However, this correlation is also affected by the inherent test-retest variability, which will depend on the applied speech material as well as the noise.

As shown in the diagram, BKB-Babble and BKB-VSE are better correlated at -2.2 dB SNR ($R^2 = 0.52$) than at 1.4 dB SNR ($R^2 = 0.42$). The same comparison applied to the effect of speech material shows that BKB-VSE and RST-VSE present comparable levels of correlation at -2.2 dB SNR ($R^2 = 0.58$) and at 1.4 dB SNR ($R^2 = 0.54$). The diagram also shows that a given change in SNR may present different degrees of subject consistency in terms of SI. For example, with BKB-Babble, there is a high correlation between the SI measured at the two SNRs ($R^2 = 0.73$). This high correlation between the two SNRs was not observed in any of the other stimuli cases ($R^2 = 0.48$ for a change in SNR in BKB-VSE and $R^2 = 0.57$ for a change in SNR in RST-VSE). The

increased correlation across SNRs that was only seen for the BKB-Babble when compared to the other two cases can likely be explained by a combination of three factors. First, SI scores were quite close to ceiling (i.e., lower variance) in the two SNRs for BKB-Babble, but not for the other cases. Second, the noise material was the same for the two SNRs. In contrast, the other two conditions had completely different noise signals for the two SNRs, as they corresponded to two completely different acoustic scenes. Hence, the lower correlation between SNRs in these cases may arise from the effect that other features of the noise signals (e.g., the modulation of the noise) have on SI, which may affect CI users differently. The reason why a slightly change in SNR in RST-VSE led to higher correlation values than in BKB-VSE may be explained by the fact that the SI scores for RST-VSE were in the linear (i.e., the steepest and most sensitive) region of the underlying psychometric function, whereas the SI scores for BKB-VSE at the high SNR were closer to ceiling (Fig. 4.4) and therefore compressed. Third, BKB-Babble will have provided the lowest test-retest variability, due to the very consistent structure and pronunciation of the BKB sentence as well as the very steady behaviour of the Babble noise.

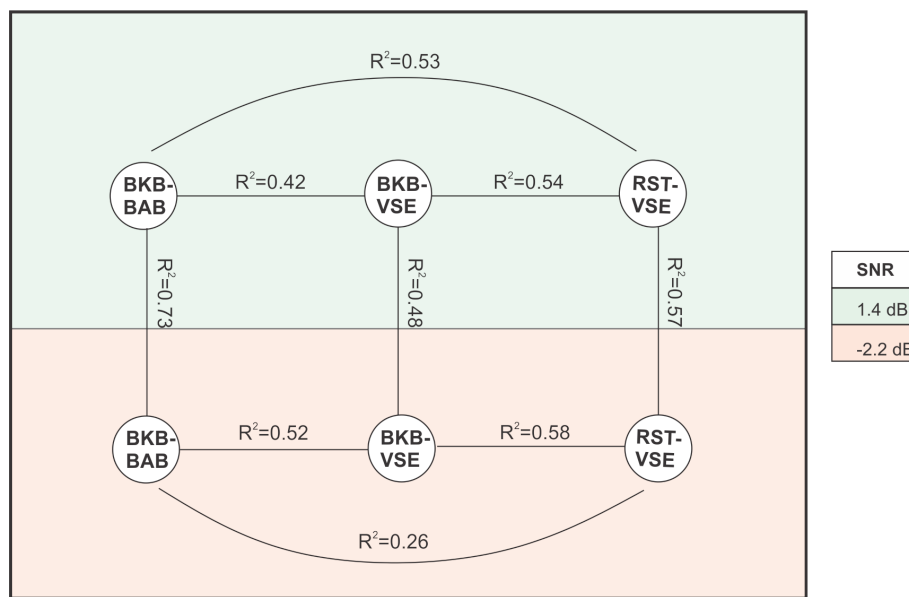


FIGURE 4.8: Diagram showing the different values of R^2 obtained between the different conditions included in the test: BKB-BABBLE (BKB-BAB in the figure), BKB-VSE and RST-VSE. The six conditions included in the test are divided in two regions according to their broadband SNR (1.4 dB, upper region and -2.2 dB, bottom region).

As shown in Fig. 4.8, a particularly low correlation is observed at the low SNR of -2.2 dB SNR between BKB-Babble and RST-VSE ($R^2 = 0.26$). However, the RST-VSE is also the condition where the highest test-retest variability is expected due to the higher variability across sentences as well as over the duration of the noise signal. Despite this potential data variability, it still seems reasonable to conclude that SI measured in the laboratory at a given SNR may not provide much information about what the person's individual SI performance would be in the real world at that same

SNR. This mismatch is probably most relevant in contexts where SI scores measured in the laboratory are used as a proxy measure of the person's SI in the real world. Although the above correlation analysis is in agreement with the common conjecture that laboratory tests may be poor predictors of real life performance and can be improved by increasing the realism of laboratory-based tests, the small number of 15 test subjects provided a limited statistical power and further data is required to confirm these results.

4.4.2 Speech intelligibility in realistic conditions

In general, the mean SI scores that were observed in all of the six more realistic (RST-VSE) conditions were rather low. Even in the office, with an SNR of 3.2 dB and a U50 of 6 dB, mean SI scores were around 60% (see Fig. 4.7). Judging from the plateau observed in the three acoustic scenes with highest SNRs (i.e., 3.2, 1.4 and -0.4 dB), it seems reasonable to suggest that the SI of the RST speech material does not increase any further, even for very positive SNRs. Measuring SI performance in quiet for the entire RST speech material with CI users might shed some light on this. Even though near 100% SI was observed for all RST sentences in quiet with young normal-hearing listeners (Kelly M. Miles et al., 2019, in preparation), SI in quiet may be significantly reduced in CI recipients. However, an alternative explanation could be that the expected increase of SI with increasing U50 values is counteracted by the increase of noise modulation, since it is known that CI recipients struggle to understand speech in modulated noise (Fu and Nogaki, 2005; Nelson et al., 2003; Qin and Oxenham, 2003). This is illustrated in Fig. 4.9, which shows the normalised modulation power spectrum of the noise and target speech signals, which was derived from simulations of an omnidirectional microphone located in the centre of the loudspeaker array by (1) applying a spectral analysis using a complex Gammatone filterbank from 125 Hz to 8 kHz (Hohmann, 2002), (2) calculating the Hilbert envelope in each frequency channel, (3) applying a modulation spectrum analysis to each envelope using a one-octave filterbank, (4) normalising the modulation spectrum in each frequency channel by the total power in each frequency channel, (5) averaging the derived modulation spectra across frequency, and (6) normalising the modulation spectrum by the maximum value observed across all considered stimuli. In Fig. 4.9, the modulation spectra are shown for the babble noise as well as for each of the six VSEs included in the study (Table 4.2). For the target speech, the average modulation spectrum of the RST and BKB-like sentences are shown, with the shaded areas denoting the 5% and 95% percentiles across the different VSEs as introduced by the different reverberation conditions.

In Fig. 4.9, the three VSEs with the highest SNRs that resulted in the same average SI scores can be compared in terms of their modulation spectra. Among them, the office is the most modulated acoustic scene, followed by the living room and the church. The office and the living room differ mostly at modulation frequencies above 3Hz but in general they present rather similar modulation spectra. In contrast, the

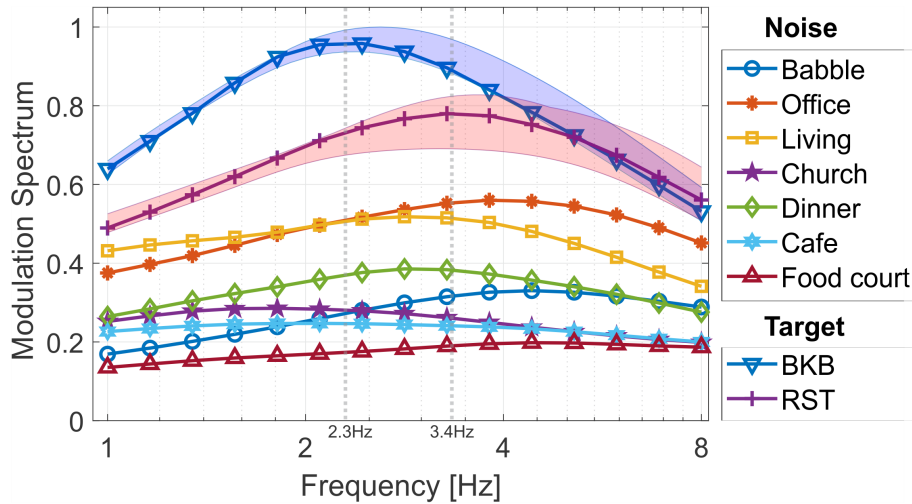


FIGURE 4.9: Normalized modulation power spectrum of each acoustic scene and target speech signals, which, for simplicity, have been grouped into BKB and RST sentences. The shaded areas denote the 5% and 95% percentiles across the different VSEs. The most prominent modulation frequencies of BKB (2.3Hz) and RST (3.4Hz) are indicated with the vertical dashed grey lines.

church presents much lower modulation values. Previous studies have shown that CI users present lower SI scores in modulated than in steady noises (Fu and Nogaki, 2005; Nelson et al., 2003; Qin and Oxenham, 2003) and it has been shown that this phenomenon can also be observed in more realistic acoustic conditions (see Chapter 3). Hence, it is reasonable to suggest that the flat modulation spectrum observed in the church may at least partly compensate for the relatively low U50 in this case. The same rationale can be applied to the living room and the office, where the U50 increases (and SI should improve) while the modulation of the noise increases as well (and SI should decrease). Hence, it is likely the case that the main reason why SI plateaus in the RST-VSE conditions at the higher SNRs is due to the noise characteristics rather than the speech material itself. However, future research should further investigate this issue, and, in particular, should measure SI for the RST sentences in CI users in quiet.

The differences between the modulation characteristics of the different VSEs may also, at least partly, explain why SI drops so abruptly from the church to the dinner party. In this case, not only does the level of the noise increase but also the modulation spectrum. This observation is also consistent with the SI scores obtained in the dinner party in relation to the ones obtained in the café. Even though the U50 in the café was 4 dB higher than in the dinner party, SI scores were only 10% higher ($t(141) = 3.207$; $p < 0.05$). Again, this may be explained by the fact that the modulation of the noise in the dinner party was substantially higher than in the café.

One last observation regarding the different modulation spectra concerns the different speech materials. As can be seen in Fig. 4.9, BKB-like sentences are more modulated than RST sentences, and they are spoken at a slower rate. As indicated

by the vertical dashed lines, for the BKB-like sentences the dominant modulation frequency is 2.3Hz, whereas for the RST sentences it is 3.4Hz. Motivated by the concept of masking in the modulation domain (Jørgensen and Dau, 2011), where the differences between the modulation spectrum of the target speech and that of the noise signals are used to predict SI, it is possible to make a rough prediction of the relative differences in expected masking for the two speech materials. For example, in the living room, the modulation spectrum evaluated at 2.3Hz or 3.4Hz (i.e., at the dominant modulation frequencies for the RST and BKB sentences respectively) are virtually the same. However, the difference between the modulation spectra at the dominant modulation frequency was 0.4 normalized units for the BKB-like sentences and 0.28 for the RST sentences. Hence, this difference in the SNR in the modulation domain may at least be one of the reasons why SI with BKB-like sentences was significantly higher than with RST sentences. However, future research will need to apply more sophisticated SI models, such as described in Jørgensen and Dau (2011), to further investigate this hypothesis.

In regards to bilateral benefit, the finding that bilateral benefit was on average 7% better than with the BPE is in agreement with previous literature. Van Hoesel (2011) reviewed 11 bilateral CI studies that measured speech performance in noise, and found good agreement across studies when the bilateral condition is compared against the better performing ear in a given listening condition (including diotic conditions). The average benefit in that case corresponds to about 1dB improvement in SNR, or equivalently 7% SI improvement assuming typical psychometric function slopes for bilateral CI users. Approximately the same benefit was obtained regardless of whether the stimuli were diotic or dichotic.

Despite the fact that bilateral benefit in the present study was shown to be quite constant across acoustic environments and subjects, a series of t-tests for individual environments showed that the benefit was not statistically significant in the church and the dinner party. While the benefit in the church was close to significant ($p = 0.051$), the results in the dinner party were clearly non-significant, a fact that was moreover consistent across the RST and the BKB speech materials ($p = 0.16$ and $p = 0.33$ respectively). In the dinner party, the U50 at the BPE was about 1dB lower than in the WPE, which should if anything, have increased the likelihood of measuring a bilateral benefit because the added ear had a better U50 (Van Hoesel, 2011, page 19). However, the results indicate that the small U50 asymmetry was of little consequence compared to the larger asymmetry of the performance of the two ears (Fig. 4.7). One speculative explanation is that the performance asymmetry between ears becomes more important at lower SNRs (such as in the dinner party), and the bilateral benefit is reduced because the added ear offers little information. It is also possible that the individual scene analyses were somewhat limited by statistical power considerations.

In the BKB-VSE data set, the bilateral benefit over the BPE in the living room was

also small and non-significant. However, in this relatively quiet scene with easy target materials, it seems likely that ceiling effects reduced the observed benefit because BPE results were already high.

4.5 Limitations and outlook

This study measured SI in noise with bilateral CI recipients using more realistic speech material, noises and SNRs. To the best knowledge of the authors, this was the first time such high level of realism was applied to a speech-in-noise test. Therefore, it was unclear how to best select the environments, what instrumental measures to use for selecting the environments (and speech levels) such that the difficulty of the different conditions varied gradually, and how difficult it would then be for the CI users. In this regard, the results revealed that the U50 would have been the more appropriate measure to select the acoustic environments (and speech levels) than the broadband SNR that was applied here. Future research may consider even more advanced SI models, as the U50 does not account for some of the attributes that are relevant to SI, such as the modulation characteristics of the noise signals.

The RST sentences used in this study have never been used before with CI recipients and need to be further evaluated. Even though an extensive evaluation with young normal-hearing listeners showed SI scores near 100% for all RST sentences in quiet, this is not clear for CI users and may, at least partly, explain why the mean SI scores plateaued at a performance level of 60%. Unfortunately, due to the major effort involved, a validation of the RST sentences in quiet was out of the scope of the present study and should be done in the future. This observation, in turn, highlights the importance of designing test paradigms with a well-adjusted trade-off between realism and control according to the question at hand. Here, the goal of the SI evaluation conducted in the RST-VSE conditions was not to disentangle the effect of the speech material from that of the SNR or from that of the modulation of the noise, but to conduct an exploratory evaluation of SI under realistic conditions. Future studies may want to constrain some of these variables at the expense of realism so that the data provides knowledge about their individual effects. With respect to test realism, it should also be considered that, in favour of better control over the signal presented to the CI speech processor and to focus on the evaluation of the involved auditory processes, most of the adaptive speech processing features (except for ASC) were turned off. Hence, future work aimed at obtaining even more ecologically-valid SI outcomes will need to include these features. Moreover the word recall tasks, like the one employed here, do not well reflect the actual tasks that are observed in the real world, where people need to comprehend what other people say or actively participate in a conversation. Future studies should therefore use other more realistic tasks, such as speech comprehension (Best et al., 2016) and communication task (Beechey, Buchholz, and Keidser, 2019).

It is often reported that limited information can be deduced from standard laboratory tests of a subject specific/individual performance in the real world. The observation here that SI measured in the most realistic conditions at -2.2 dB SNR correlated particularly poorly with the least realistic (standard) condition may suggest that this more realistic condition better reflects real life performance. However, to support these conclusions it is required that far more subjects are tested to improve the statistical power and direct comparisons should be made with field-studies that measure performance with the same subjects in the real world. Future research in this regard should also consider the addition of visual cues in the laboratory.

As a final remark, it should be noted that the effect of realism on SI outcomes, as it was explored in this study, only compared three very specific test conditions (i.e., BKB-Babble, BKB-VSE, and RST-VSE) and therefore, it is unclear how far the results and conclusions can be generalised to other test conditions. For instance, the comparison did not include the RST-Babble combination, and the different sentence materials did not only differ in their level of realism, but involved also two different female talkers. These and other potential differences between conditions may have not been solely attributed to their level of realism and may have therefore influenced SI outcomes. However, it is difficult to draw a clear line between factors that are strictly related to realism from those that are not. Hence, experiments need to be designed in the future that can better control these factors while still providing a high level of realism.

4.6 Conclusions

The study had two main goals. The first goal was to explore the effect of the realism of the target speech and noise on the ability of bilateral cochlear implant users to understand speech in noisy conditions. To achieve this goal, three different stimulus conditions were compared at two different SNRs (i.e., at 1.4 dB and -2.2 dB): (1) anechoic BKB-like sentences presented in babble noise, (2) reverberant BKB-like sentences presented in three-dimensional recordings of actual noisy situations, and (3) a variation of the second test where the sentences were taken from natural effortful conversations. All the tests were conducted by means of a spherical loudspeaker array located inside an anechoic chamber. Speech intelligibility results were consistently highest in the first test (the least realistic) and lowest in the third test (the most realistic). These results indicate that participants could more easily deal with babble noise than with more realistic noisy situations, and that they could understand more easily BKB-like sentences than more realistic speech. The effect of the more realistic noise on SI was fully explained by the U50 of the acoustic signals that arrived at the directional input of the speech processor. However, the U50 was not able to explain the effect of the more realistic speech material. A correlation analysis between the different test conditions revealed a particularly low correlation between the most realistic and the least realistic (or standard) test condition for the low SNR of -2.2

dB. This pronounced intra-subject variability may suggest that the more realistic test assessed different auditory functions than the standard test, and may therefore better represent the individual SI ability observed in the real world. However, further studies with far more subjects are required to substantiate this conclusion. The second goal of this study was to evaluate the ability of CI users to understand speech over a range of different realistic conditions. For this purpose, the more realistic speech material was presented in six realistic acoustic scenes at realistic SNRs and speech intelligibility was measured both unilaterally and bilaterally. Speech intelligibility scores increased with increasing SNR, but plateaued over the three highest SNRs with mean scores not exceeding 60%. Even though it could not be ruled out that this plateau effect was caused by the speech material, evidence was provided that an increased amount of noise modulation may have also counteracted the expected increase in SI within increasing SNR. Future work will need to further investigate these factors. Bilateral hearing provided a consistent 7% benefit over unilateral speech understanding in the better-performing ear, which was significant in most of the acoustic environments.

Appendices

4.A Biographic data of participants

TABLE 4.A.1: Biographic data of the participants. ID refers to the identifier of each subject employed throughout this study. Gender is either female (F) or male (M), age is the age of the participant at the time of running the tests, processor is the speech processor model, implant is the type of implant, age at implantation is the age of the participant at the time of the surgery, hearing loss is the aetiology of hearing loss, and the last column provides information about the subjects' aided SI scores (i.e. with hearing aids) obtained in quiet conditions with CUNY sentences at 65 dB SPL right before implantation.

ID	Gender	Age	Side	Processor	Implant	Age at implantation	Hearing loss	SI with HA
1	F	29	L	N6	CI512	20	Autoimmune disease	NA
			R	N6	CI24RE (CA)	21	Autoimmune disease	24%
2	M	62	L	N7	CI422 (SRA)	56	Meniere's Disease	NA
			R	N7	CI522	59	Meniere's Disease	94%
3	F	52	L	N7	CI24RE (ST)	56	Familial progressive	17%
			R	N7	CI24R (CS)	59	Familial progressive	20%
4	F	61	L	N7	Ci24M	59	Progressive unknown	NA
			R	N7	CI24RE (CA)	47	Progressive unknown	7%
5	F	61	L	N7	CI522	59	Progressive unknown	2%
			R	N7	CI24RE (ST)	47	Progressive unknown	64%
6	F	45	L	N7	CI512	43	unknown	NA
			R	N7	CI24RE (CA)	39	unknown	NA
7	F	44	L	N7	CI24RE (ST)	33	Ototoxic medication	NA
			R	N7	CI24R (CS)	26	Ototoxic medication	NA
8	F	61	L	N7	CI24RE (CA)	51	acquired progressive idiopathic	NA
			R	N7	CI512	39	acquired progressive idiopathic	NA
9	M	65	L	N7	CI512_II	63	Meningitis	NA
			R	N7	CI512_II	63	Meningitis	NA
10	M	66	L	N7	CI24R (ST)	45	meningococcal meningitis	NA
			R	N7	CI24RE (ST)	32	meningococcal meningitis	NA
11	M	60	L	N7	CI512	51	Radiation for Nasopharyngeal cancer	0%
			R	N7	CI512	50	Radiation for Nasopharyngeal cancer	0%
12	F	63	L	N7	CI422 (SRA)	58	Progressive unknown	98%
			R	N7	CI522	62	Progressive unknown	50%
13	F	61	L	N7	CI24RE (ST)	47	Genetic, familial	32%
			R	N7	CI24RE (CA)	51	Genetic, familial	31%
14	F	48	L	N7	CI422	43	Sudden idiopathic	0%
			R	N7	CI422	43	Sudden idiopathic	0%
15	M	57	L	N7	CI522	57	Ushers syndrome	94%
			R	N7	CI522	56	Ushers syndrome	90%

4.B Speech Reception Thresholds

TABLE 4.B.1: Averaged SRTs measured at each ear of the 15 participants with BKB-like sentences and collocated (3x) four talker babble

ID	SRT left (dB)	SRT right (dB)	BPE (or preferred ear)
1	1.4	1.2	L
2	7	2.2	R
3	1.2	1.5	R
4	3.8	-0.6	R
5	3.4	3.7	R
6	3.3	0.3	R
7	1.2	0	R
8	0.7	2	L
9	3.6	3.4	L
10	6.3	12.4	L
11	0.3	2.8	L
12	0.9	3.3	L
13	4	4.3	R
14	0.4	0.7	L
15	0.8	-0.3	L

Chapter 5

Final considerations

5.1 Discussion

This thesis investigated the effect of realistic noisy and reverberant conditions on speech intelligibility (SI) outcomes in cochlear implant (CI) users. This was achieved by reproducing, in the laboratory, the sound field of realistic acoustic scenes that CI users may encounter in their daily lives. The setup and methods required to achieve such a level of realism in the laboratory relied on the concept of Higher Order Ambisonics and multi-channel loudspeaker-based reproduction (Oreinos, 2015b). With this framework in place, reasonably accurate reproductions of sound fields can be achieved, which enables participants to use auditory cues in a comparable manner as they would in real life.

Although one of main contributions of this thesis to “realistic testing” lies in the three-dimensional reproduction of sound fields, one aspect that was here incorporated and shown to be ecologically relevant was the realism of the scenes to be reproduced. Although this may sound trivial, the results obtained here showed that the main reason why most of the previous studies found such a detrimental effect of reverberation on SI in CI users was due to the choice of the room employed during the listening tests rather than the use of simple playback methods (Hu and Kokkinakis, 2014; Kokkinakis, Hazrati, and Loizou, 2011; Kokkinakis and Loizou, 2011). The clearest example of such odd reverberation conditions consists of small room volumes with long reverberation times (RTs), which are rarely encountered in real life and very detrimental to SI. In contrast, the present study reproduced real physical rooms and made an effort to cover a wide range of distances and acoustically distinct rooms in which everyday communication frequently occurs.

The use of very challenging reverberation conditions, as the ones employed in previous studies, could be justified by the need to ensure that SI outcomes do not suffer from ceiling effects, which would certainly impede any interesting observation of reverberation effects. While this argument is valid, it raises a number of concerns. First, the relevance of such experimental conditions in terms of real-life performance is questionable. For example, signal processing techniques designed to mitigate the effects of such reverberation conditions would be required as often as people encounter those situations. Second, if the variety of real-world reverberant spaces is

not accounted for, the methods by which challenging reverberation conditions are achieved may have an impact on the conclusions drawn. This is what has been observed here in regards to previous studies, as the effect of reverberation on SI was uniquely described by the RT (Desmond, Collins, and Throckmorton, 2014; Hazrati and Loizou, 2012; Hu and Kokkinakis, 2014; Kokkinakis and Loizou, 2011; Kokkinakis, Hazrati, and Loizou, 2011). In contrast, the present study was able to show that the RT cannot be considered to be a unique descriptor of SI, because this measure does not provide any information about the contribution of the direct sound relative to the pressure of the reverberant field (i.e., the DRR). As previously discussed, this misconception likely arises from the use of simplistic methods where different reverberant conditions are obtained by altering the absorption material of a single room. Because increasing RTs lead in these cases to decreasing DRRs, SI decreases monotonically with increasing RTs. However, the drop in SI performance in these cases cannot be attributed to the effect of the increasing RT alone, because other factors such as the decrease in the DRR (or the frequency response as briefly discussed in Sec. 3.4.1.2) have or may have an impact. Should the researcher be interested in the effect of the RT alone, all the other potential factors should be kept constant by, for example, testing at the critical distance and equalising the spectrum of the different stimuli.¹ Another concern arising from such testing methods can be seen as a side-effect of the wrong assumption that the RT is a unique descriptor of SI. The RT is a measure that reflects reasonably well the perceived reverberation of a room. Hence, the link between RT, SI and perceived reverberation may be misleading in situations where for example a misinformed clinician advises a customer to avoid reverberant spaces because (they have very long RTs and hence) they are very detrimental to SI. The person may therefore be unnecessarily discouraged to go to certain places and participate in social activities. Although the RT can be roughly associated with the perceived reverberation, SI certainly cannot. Therefore, when talking in terms of SI, it is necessary to acknowledge the other important *dimension* of reverberation, the DRR, which is especially relevant in CI users.

Avoiding floor and ceiling effects is not new, and it is in fact the principle upon which the use of the Speech Reception Threshold (SRT) is based. Without denying the usefulness of such a measure, several concerns can also be raised in this case. First, as mentioned several times in this document, paying attention to the acoustic conditions in which a given person understands half of what is being said is not relevant in terms of real life performance, because people very rarely communicate under such conditions. The second concern is related to the fact that the SRT corresponds to the point in the performance intensity function at which the sensitivity is highest. For example, assuming a maximum slope of 12%/dB, a speech processing technique providing 1 dB SRT benefit over another processing technique implies that the maximum benefit obtained is 12%. That benefit would be obtained in situations

¹Ironically, such a test would probably lead to monotonically decreasing SI scores with increasing RT values, as suggested by Eq. (3.4)

where the person understands only half of what is being said. Because these situations rarely occur, the translation from the SRT to SI is misleading and overestimates SI.

Regardless of how often a person may encounter the rooms employed in the laboratory, the use of accurate descriptors of SI in reverberation with CI users makes it easier to compare SI scores obtained across different research laboratories. Understanding the different reverberant conditions employed in previous studies was here found to be particularly challenging given the fact that the RT reported did not provide an insight about SI. In contrast, the present study showed that the U50 is a far more suitable measure to predict the effect of reverberation on SI, a fact that makes it possible to compare the results of future studies against the ones obtained here (irrespective of the talker-to-listener distance, the room volume and the RT). Moreover, the U50 was also be very useful to understand the individual effects of the RT and the DRR, by means of which the effect of basic parameters such as the talker-to-listener distance, the room volume or the RT itself could be well understood. This not only provided a good insight of how these parameters may affect SI under realistic reverberant conditions, but also a general framework that enabled a qualitative comparison across studies thereby explaining the apparent discrepancies observed between them. In fact, the use of more complete metrics such as the U50 or the STI (Whitmal and Poissant, 2009; Poissant, Whitmal, and Freyman, 2006), where different experimental methods, conditions and setups can be jointly analysed, enables different results to point in the same direction, rather than diverging from one another.

In line with the overarching goal of gaining insight about the difficulties faced by CI users in the real world, STI ratings between people with normal hearing (NH) and CI users were compared. The comparison clearly showed that CI users struggle in situations where NH listeners have no problem at all. For example, SI for NH listeners in an environment rated as "fair" ranges within approximately 70% and 98%. According to the results obtained here, the same environment in CI users would lead to average SI scores between 6% and 46%. Hence, although reverberation was not here found to be as detrimental to SI as in other studies, reverberation is a problem for CI users. As discussed here and in many other studies, the difficulties of CI users in reverberation stem from SI with CIs relying heavily on the envelope of the speech signal, which becomes more flat with the effect of reverberation. Apart from solutions based on the pre-processing stage of the speech processor like beamformers or compressors, future solutions may want to evaluate the mechanisms employed by NH listeners for whom SI in reverberation is extremely robust (Nabelek and Pickett, 1974).

Due to the large differences observed between CI users and NH listeners, it was here considered unlikely that SI criteria of acoustically treated venues will be based on the necessities of people who rely on CIs. Rather, solutions based on FM or other wireless systems should be more often adopted in the future to ease SI in acoustically

treated venues. In any case, the observation that the U50 was a suitable predictor of the effect of reverberation on SI with CI users can prove useful in the pre-evaluation of any of the solutions adopted in terms of SI improvement. In line with this, the U50 could also be used in the improvement of speech enhancement methods of the person's speech processor like for example de-reverberation algorithms. However, SI in reverberation was here assessed without any speech processor directivity (i.e. omnidirectional) and most adaptive speech processing features were turned off. Hence, for a clearer statement of the problem, the needs and the subsequent proposed solutions, future research should evaluate the effect of reverberation with all these advanced features, including the microphone beamformer, enabled.

The observation that modulated noises hinder SI more than steady noises is not new. Several studies have observed this phenomenon in the laboratory by means of gated noises presented at different rates (e.g. Nelson et al., 2003; Fu and Nogaki, 2005). The difference between those studies and the present one is that here it was possible to confirm that the effect of the modulation of the noise is an ecologically-relevant variable, as it had a systematic effect on SI when evaluated under realistic conditions. Interestingly, the observations made here would not have been possible without the knowledge provided by previous studies (e.g. Nelson et al., 2003; Fu and Nogaki, 2005), which were conducted with highly controlled experimental methods. This is a clear example showing that highly realistic tests and highly controlled tests are not mutually exclusive but rather, attempt to answer different questions. While the focus of highly controlled tests is to learn about the auditory processes involved, in this case, in speech understanding, highly realistic tests are more directed towards the real-life performance. While highly controlled tests are crucial to learn from the data, highly realistic tests are crucial to determine how relevant is the data. Although it may seem as if these two approaches cannot be combined, the approach adopted in this thesis tried to find a good trade-off between control and realism whenever they were in conflict. In the beginning of this thesis, the effect of reverberation on SI in both quiet and noisy conditions was extremely unclear. Hence, the test paradigm designed to learn what the actual effect of reverberation was had to trade off some realism thereby making sure that the question at hand would be answered. Although the original noise signals and the testing SNRs were deliberately kept constant across conditions to ensure that they differed exclusively due to the time smearing effect of reverberation (in favour of control), the realism was still maximised by reproducing the sound field of actual and relevant rooms (in favour of realism). Hence, the studies included in this thesis attempted to apply a high level of realism while making sure that some categorical statements could be made from the data.

The analysis based on the U50 made it possible to show that the main ecologically relevant factors affecting SI in rooms were the ratio between early and late reflections of the target speech (i.e., the C50), the level of the noise and the modulation of the noise. Establishing the U50 as a SI measure, along with the modulation of the noise, was an important milestone in the present thesis, as it eased the analyses of the data

obtained in the subsequent study, which corresponded to more realistic (less controlled) conditions.

Three-dimensional reproduction of sound fields was not the only contribution of this thesis to “realistic testing”. As has been previously discussed here and elsewhere (e.g. Cord et al., 2007) there are multiple dimensions over which laboratory-based assessment of SI can become more ecologically relevant. As already mentioned and described in previous sections, the last study included in this thesis incorporated the concept of realistic SNR, realistic noise scenes and realistic speech test (RST) material. In order to investigate the effect of test realism on SI scores, three different levels of test realism were included in the design of the test. The analysis of the data revealed that the most realistic conditions (RST-VSE) were consistently the most difficult in terms of SI whereas the most artificial conditions (BKB-Babble) were the easiest. Although it is here acknowledged that the conditions chosen here were just an example among many possibilities, the results are in agreement with the commonly reported mismatch between performance in the laboratory and the real world (e.g. Cord et al., 2007). Participants were also quite grateful in regards to the relevance of the speech-in-noise tests. In regards to the RST speech material, one of the participants said *“For the first time I feel like there is meaning behind a speech test”*. This is not extremely surprising considering how different speech of, for instance, BKB-like sentences is from that encountered in the real world. And as it was observed here, the differences between them had a great impact on SI scores. The disappointment of a CI user finding out that their real-world SI performance is not anywhere close to what had been suggested in the laboratory can probably be analogised to the disappointment of a person learning a new language and finding out that they only understand the overly articulated speech of their teacher.

The realism of the noise material was also praised by the subjects. One of them, after being tested in a living room with kitchen noise from the back, said *“This was so real I could even smell the bacon”*. This indicates that the scenes employed here better represented the listening experience observed in the real world. This, along with the poor correlation values obtained between the SI scores of standard (i.e., BKB-Babble) and more realistic (RST-VSE) tests, may indicate that the latter assessed different auditory functions than the former. Unfortunately, these are big claims that would require far more subjects and factorial experimental designs.

5.2 Outlook and limitations

Although most of the limitations were already highlighted in the specific chapters, this section provides a broader overview of them.

The first limitation of this thesis concerns the advanced pre-processing features that were disabled during the listening tests: Automatic Dynamic Range Optimisation (ADRO), SNR-NR, Spatial-NR, SCAN, WhisperTM, WNR and, in the first two

studies, the microphone directivity. While this decision enabled a systematic investigation of the signals reaching the BTE processors, it may have moved the performance away from that obtained in real life. For example, in the study concerned with the effect of reverberation, higher U50 values would have been obtained if the microphone directivity had been enabled. Future research aiming at measuring more ecologically-valid speech intelligibility outcomes should conduct tests in which all these features are enabled. Moreover, it is questionable whether the U50 could provide reliable information about the benefit provided by these adaptive features, as it is a time-integrated measure.

The second limitation concerns the stimuli. Throughout the present study, the limitations in relation to the stimuli employed have been highlighted several times. For example, in the first study, target speech was always presented at 60 dB SPL regardless of the talker-to-listener distance. Future studies concerned with the effect of distance on audibility should consider more realistic stimuli whose sound pressure levels decay with distance. If that is the case, ADRO would have been to be enabled, as it would likely have a strong impact on speech intelligibility outcomes. Likewise, the noise layout employed in Chapters 2 and 3 was arbitrarily designed as a good representative of a typical noise environment but it is currently not clear whether a different noise layout would have led to different results. Moreover, the tests were conducted at a fixed but subject-dependent SNR. Future studies following up on the validity of the U50 will likely benefit from testing speech intelligibility under different SNRs and comparing subject performance at the same SNRs. Another limitation in regards to the stimuli employed was found in the last study, where it was acknowledged that the different levels of realism corresponded to specific examples and that many other possibilities and testing SNRs could have been chosen instead.

The third limitation of this thesis is related to the novelty of the research. This was the first time such a level of realism was applied to speech intelligibility tests with CI users. Hence, it was hard to select the different environments and to estimate how difficult speech understanding would be. Moreover, this was the first time the RST speech material was used with CI users. It is unclear whether CI recipients can actually reach 100% SI scores with this material when tested in quiet conditions. Hence, it is hoped that in the future, new evidence and knowledge obtained from different research institutions will progressively be added together in such a way that future experimental tests will be less exploratory, easier to design and easier to analyse.

The fourth limitation is related to the previous limitation and concerns the trade-off between realism and control. Although it was here the intention to apply highly realistic conditions while making sure the questions at hand would be answered, the last study left one question open. In particular, the plateau of the RST-VSE SI data observed at the three highest SNRs could not be categorically attributed to a specific factor because it could be the consequence of two uncontrolled factors. One possibility was that the RST speech material was so challenging for CI users that SI

would not have been better even at more positive SNRs. The second possibility was that the increasing noise modulation observed at increasing SNR prevented CI users from exhibiting increasing SI scores. Hence, future research will have to clarify this issue before the speech material is used in noisy conditions.

Another limitation concerns the accuracy of the sound field reproduction. As explained in Sec. A, an accurate reproduction of the real sound field (i.e., that of the actual noisy acoustic scene) is only possible up to a certain frequency (and area) that depends on the number of transducers (microphones and loudspeakers) of the system. The current study applies two strategies to increase this frequency limit from that obtained with the classic HOA formulation. The first strategy consists of applying the concept of Mixed Order Ambisonics (MOA) whereby the horizontal plane has a higher density of loudspeakers thereby extending the usable frequency range in the horizontal plane, where directional hearing is most acute. A thorough assessment of the sound field reproduction errors conducted in the loudspeaker array employed here revealed two important characteristics of the system. First, the reproduction error is higher at the ear contralateral to the virtual source (Oreinos and Buchholz, 2016). Second, the reproduction of reverberant sound sources exhibits lower errors than that of anechoic sources (Oreinos and Buchholz, 2016). Following up from these observations, the second strategy applied in the current study aims at extending the frequency range of the direct sound (that is, the anechoic part) of reverberant sound sources. This strategy was here applied to all the sound sources based on RIRs by enforcing the direct sound to be reproduced by a single loudspeaker. This strategy enabled faithful reproduction of the direct sound up to the highest frequency of the signal. However, the reproduction of recorded acoustic scenes (namely, those employed in Chapter 4) still suffered from the highlighted limitations and hence, future research will have to provide new ideas to overcome them.

Moreover, future research is needed to implement realistic testing in audiological clinics, as it is currently unrealistic to envision solutions based on spherical arrays of multiple loudspeakers that, moreover, need to be installed in an anechoic chamber. Binaural audio presented via headphones is currently the easiest solution, although is not free of limitations. To name a few difficulties, individualised spatial cues are difficult to convey and participants cannot use head movement to improve localization (unless head tracking systems are incorporated). Although the future of realistic hearing tests in the clinic is highly uncertain, potential future research may want to consider the possibility of designing loudspeaker arrays with a small radius comprising smaller loudspeakers located very close together thereby minimizing spatial aliasing and allowing the reproduction of intended directivities up to higher frequencies. Moreover, the employed loudspeakers could present dipole directivity characteristics to minimize the sound energy radiated outwards thereby reducing the effect of the room on the acoustic signal. However, solutions based on this approach would likely not be based on HOA but rather, on numerical inverse problems, as the assumptions of plane waves and free field would not be fulfilled.

5.3 Conclusions

This thesis represents a step forward in the evaluation of speech intelligibility of cochlear implant users considering realistic sound environments.

The first part of this thesis was concerned with the effect of realistic reverberation on speech intelligibility in quiet and noisy conditions. The results in quiet indicated that at conversational distances, speech intelligibility in most of the rooms considered was not compromised. The exception was found to be a small reflective room, where speech intelligibility was exceptionally low. The U50 was seen to be a suitable predictor of speech intelligibility, able to predict the individual's performance in all the 17 reverberant conditions considered. Moreover, the U50 was shown to be a quite interpretable measure, which provided an insight about the basic room parameters affecting speech intelligibility in reverberant conditions. This in turn helped explain why other studies found such a strong effect of reverberation on speech intelligibility and why the methods followed led the researchers to conclude that the reverberation time of a room provides enough information to predict speech intelligibility. However, the U50 did not perform as well in noise, where the effect of the modulation of the noise was seen to have an impact on speech intelligibility. In particular, the smearing effect of reverberation was seen to flatten the envelope of the noise signals making them less effective maskers. This beneficial effect of reverberation on the noise signal contrasted with its detrimental effect on the target signal in such a way that the most favourable conditions in quiet did not correspond to the most favourable conditions in noise. More specifically, when considering noise and quiet conditions together, the best reverberant conditions were shown to be short talker-to-listener distances and large volumes or smaller volumes with some reverberation. In view of the strong dependence of speech intelligibility of CI users on the modulation characteristics of noise and target speech signals, future speech intelligibility models could benefit from predictions obtained in the modulation domain.

The second part of the thesis had two goals. The first goal was to evaluate the effect of test realism on speech intelligibility outcomes. Among the three levels of test realism compared, speech intelligibility was highest in the least realistic test condition (i.e., standard BKB-like sentences with babble noise) and lowest in the most realistic condition (realistic speech and noise materials). The effect of speech realism was shown to be more noticeable at the higher tested SNR (1.4 dB) while the effect of noise realism was shown to be more relevant at the lower tested SNR (-2.2 dB). However, while the effect of the more realistic noise could be fully explained by the U50, the effect of the more realistic speech material was beyond what the U50 could explain. A correlation analysis revealed a low correlation between the scores obtained in the most and the least realistic conditions at the lowest SNR under test (-2.2 dB). Future research is required to determine whether this low correlation is due to the fact that the more realistic test assessed different auditory functions from the

standard test. The second goal was to measure speech intelligibility in highly realistic conditions both unilaterally and bilaterally. Speech intelligibility increased with increasing SNRs but plateaued over the three highest SNRs. As was the case with the evaluation of test realism, the U50 could not entirely explain the results observed. In this case, although one of the potential explanations of the plateau observed was the speech material itself, evidence was provided to suggest that the noise modulation of the different acoustic scenes may have been the main factor. With respect to bilateral benefit, bilateral performance was on average 7% better than the one attained with the better performing ear, a finding in agreement with previous studies.

In summary, the present thesis highlights the importance of conducting realistic testing. The more ecologically-valid outcomes obtained in the present study contrasted with existing findings in two different ways. First, the impact of realistic reverberation on speech intelligibility is not as detrimental as previously suggested. Second, speech-in-noise outcomes obtained in standard conditions do not necessarily correlate with those obtained in more realistic conditions. Hence, although challenges exist with *bringing the real-world into the laboratory*, it does help to shed light on the challenges that CI users face in the real world and to set the way forward to define strategies specifically devised to overcome them.

Appendix A

Higher Order Ambisonics

Higher Order Ambisonics (HOA) is based on the spherical harmonic decomposition of a source-free sound field, which arises from solving the wave equation in spherical coordinates.

The notation adopted in this study uses real-valued spherical harmonic functions. Likewise, the sign convention of the spherical coordinates are as follows: the azimuth angle θ increases counter-clockwise as observed from the positive z half-space and the elevation angle δ increases towards positive z values. The same sign convention and notation has been widely used in the relevant literature (see e.g. Daniel, 2001 and Bertet, Daniel, and Moreau, 2006).

The pressure at a point $\mathbf{r}=(r,\theta,\phi)$ inside a source-free region can be expressed as

$$p(kr, \theta, \phi) = \sum_{m=0}^{+\infty} i^m j_m(kr) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \phi) \quad (\text{A.1})$$

where k is the wavenumber, i is the imaginary unit, $j_m(kr)$ is the spherical Bessel function of degree m , B_{mn}^{σ} are the HOA components and $Y_{mn}^{\sigma}(\theta, \phi)$ is the spherical harmonic function of degree m and order n , defined as

$$Y_{mn}^{\sigma}(\theta, \phi) = \sqrt{(2m+1)(2-\delta_{0,n}) \frac{(m-n)!}{(m+n)!}} P_{mn}(\sin \phi) \times \begin{cases} \cos(n\phi), & \text{if } \sigma = +1 \\ \sin(n\phi), & \text{if } \sigma = -1 \text{ (ignored if } n = 0), \end{cases} \quad (\text{A.2})$$

where $\delta_{0,n}$ is the Kronecker delta and P_{mn} are the associated Legendre functions.

A.1 Sound field encoding

The process of finding the HOA components of a sound field is commonly referred to as HOA encoding. Consider a set of Q infinitesimally small microphones located on the surface of a rigid sphere of radius R . The sound pressure at the location of the microphone q can be expressed as (Bertet, Daniel, and Moreau, 2006):

$$p_R(kR, \theta_q, \phi_q) = \sum_{m=0}^{+\infty} W_m(kR) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta_q, \phi_q). \quad (\text{A.3})$$

The term $W_m(kR)$ is here referred to as the filter function and it is expressed as:

$$W_m(kR) = i^m \left(j_m(kR) - \frac{j'_m(kR)}{h'_m(kR)} h_m(kR) \right), \quad (\text{A.4})$$

where h_m is the spherical Hankel function, and h'_m and j'_m are the derivatives with respect to r of the spherical Hankel function and the spherical Bessel function respectively.

In practice, the infinite series in Eq. (A.3) is truncated to an order M , which is normally referred to as the order of the HOA system. Such a truncation leads to a total number of HOA components B_{mn}^σ equal to $(M+1)^2$. The criterion to select a value of M is related to the number of sensors Q comprising the microphone array. Roughly speaking, a perfect estimate of the HOA components can be achieved only if the number of microphones Q is higher or equal to the number of HOA components to be estimated, i.e. $Q \geq (M+1)^2$. For instance, an array of 64 microphones is able to estimate the HOA components of a sound field up to a 7th order. The maximum order for an accurate estimation of the ambisonic components poses either a frequency or a distance limit from which spatial aliasing errors occur. As a rule of thumb, the order that keeps the reproduction errors within a certain range (normalised mean square error lower than -14 dB) can be expressed as $M = \lceil kR_r \rceil$, where R_r is the radius of reproduction and $\lceil \cdot \rceil$ denotes the higher nearest integer (Bertet, Daniel, and Moreau, 2006). For instance, reproduction up to a 7th order over a spheric region of 0.1m radius would lead to errors higher than -14 dB from 3.8 kHz onwards. If for instance the reproduction is intended over a sphere of 0.2m, the frequency limit becomes 1.9 kHz.

After truncation, Eq. (A.3) can be expressed in matrix notation as:

$$\mathbf{Y} \text{diag}[\mathbf{W}(kR)] \mathbf{b} = \mathbf{p}_R, \quad (\text{A.5})$$

where \mathbf{Y} is a $[Q \times (M+1)^2]$ matrix that contains $(M+1)^2$ spherical harmonic functions $Y_{mn}^\sigma(\theta, \phi)$ sampled at Q different microphone positions (r, θ_q, ϕ_q) , $\text{diag}[\mathbf{W}(kR)]$ is an $[(M+1)^2 \times (M+1)^2]$ diagonal matrix that contains the filter functions, and \mathbf{b} is an $[(M+1)^2 \times 1]$ vector comprising the HOA components to be estimated. Note that all matrices are assumed to be full rank, unless otherwise stated. Refer to Sec. A.4 for further details.

The least-norm solution of Eq.(A.5) $\hat{\mathbf{b}}$ is:

$$\hat{\mathbf{b}} = \text{diag}[\mathbf{W}(kR)]^{-1} \text{pinv}(\mathbf{Y}) \mathbf{p}_R = \mathbf{E} \mathbf{p}_R, \quad (\text{A.6})$$

where $\text{pinv}(\mathbf{Y})$ is the pseudoinverse of \mathbf{Y} , which, if $Q > (M + 1)^2$ (overdetermined system of equations) is defined as $(\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T$, where the superscript $(\cdot)^T$ denotes the matrix transpose. The matrix $\mathbf{E} [(M + 1)^2 \times Q]$ denotes the encoding matrix, which enables the conversion from the pressure at the microphone positions to the HOA components of the sound field.

A.2 Shape matching

The pressure at a point $\mathbf{r}=(r,\theta,\phi)$ due to a plane wave arriving from a direction (θ_l, ϕ_l) can be expressed as (Bertet, Daniel, and Moreau, 2006)

$$p(kr, \theta, \phi) = \sum_{m=0}^{+\infty} i^m j_m(kr) \sum_{n=0}^m \sum_{\sigma=\pm 1} Y_{mn}^\sigma(\theta_l, \phi_l) Y_{mn}^\sigma(\theta, \phi), \quad (\text{A.7})$$

which, by comparison with Eq.(A.1), clearly shows that the ambisonic components of a single plane wave with known direction can be expressed as

$$B_{mn}^\sigma = Y_{mn}^\sigma(\theta_l, \phi_l). \quad (\text{A.8})$$

Eq. A.8 is useful only to encode a single plane wave arriving from a known direction; for a more general case, where sounds arriving from arbitrary directions are picked up by the microphone array, Eq. (A.6) must be used instead. However, Eq. (A.6) is not very useful in practice because the frequency response, the actual positions and the potential diffraction effects of the flush-mounted microphones comprising the array are not accounted for. Besides, Eq. (A.6) relies on the assumption that the hard-sphere model accurately expresses the pressure on the microphones comprising the array (Chapter 2, Oreinos, 2015a).

Shape Matching (Bertet, Daniel, and Moreau, 2006, Equation 46) is a technique that defines an encoding matrix \mathbf{E} that can be applied to any arbitrary plane wave field while accounting for the frequency response of the microphones as well as any positioning errors. The technique uses the knowledge that for a plane wave of known direction the true ambisonic components are known (Eq. A.8). It is then possible to define a matrix of true ambisonic components \mathbf{C} of size $[(M + 1)^2 \times L]$, spatially sampled according to a spherical grid of L directions under consideration.

Although \mathbf{C} is referred to here as true ambisonic components, it is important to highlight that they come from sampling the spherical harmonic functions at a finite number of positions/directions. In our case, we use a spherical array of (previously equalized) 41 loudspeakers, giving rise to a grid of $L = 41$ plane waves. Any linear combination \mathbf{b} of the sampled, true ambisonic components (i.e., spherical harmonics sampled at the loudspeaker positions) can be expressed as:

$$\mathbf{b} = \mathbf{C}\mathbf{s}, \quad (\text{A.9})$$

where \mathbf{s} is the loudspeaker driving signals. The pressure at the microphone positions can be expressed as:

$$\mathbf{p}_R = \mathbf{H}\mathbf{s}, \quad (\text{A.10})$$

where the matrix \mathbf{H} ($[Q \times L]$) contains all the complex transfer functions between the pressure at the microphones and the loudspeaker input signal. The goal of Shape Matching is to obtain an encoding matrix \mathbf{E} such that

$$\mathbf{b} = \mathbf{E}\mathbf{p}_R. \quad (\text{A.11})$$

By inserting Eq. A.9 and Eq. A.10 into Eq. A.11, it can be seen that

$$\mathbf{E}\mathbf{H} = \mathbf{C}. \quad (\text{A.12})$$

The encoding matrix will then be:

$$\mathbf{E} = \mathbf{C}\text{pinv}(\mathbf{H}). \quad (\text{A.13})$$

If $Q > L$, Eq. A.13 is an overdetermined system of equations, leading to the expression:

$$\mathbf{E} = \mathbf{C}\mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}, \quad (\text{A.14})$$

where the superscript $(\cdot)^H$ denotes the Hermitian transpose.

A.3 Sound field decoding

HOA decoding normally refers to the process of finding the driving signals of a loudspeaker array from the knowledge of a HOA-encoded sound field. Because HOA assumes that both the true and the reproduced sound fields consist of a superposition of plane waves, the decoding process can be formulated as finding the amplitudes of a finite set of plane waves that better approximate the true plane wave field. Equating the loudspeakers' plane wave superposition to a true plane wave field represented by a wave vector \mathbf{k} yields:

$$\sum_{l=1}^L s_l \sum_{m=0}^M \sum_{n=0}^m \sum_{\sigma=\pm 1} Y_{mn}^{\sigma}(\theta_l, \phi_l) = \sum_{m=0}^M \sum_{n=0}^m \sum_{\sigma=\pm 1} Y_{mn}^{\sigma}(\theta_k, \phi_k), \quad (\text{A.15})$$

where s_l is the driving signal of the loudspeaker l . Note that Eq. A.15 can be generalised to the case of an arbitrary field expressed as a linear combination of $Y_{mn}^{\sigma}(\theta_k, \phi_k)$ obtained during the encoding process of the sound field (Eq. A.11). Note also that Eq. A.9 ($\mathbf{b} = \mathbf{C}\mathbf{s}$) corresponds to the formulation of the decoding process (the generalisation of Eq. A.15) expressed in matrix form. Because the goal in this case is to obtain the driving signals of the loudspeakers, solving Eq. A.9 for \mathbf{s} results in

$$\hat{\mathbf{s}} = \text{pinv}(\mathbf{C})\mathbf{b}. \quad (\text{A.16})$$

where $\text{pinv}(\mathbf{C})$ is normally referred to as the decoding matrix \mathbf{D} . If $L \geq (M+1)^2$, the least-squares solution to the underdetermined system of equations is expressed as

$$\hat{\mathbf{s}} = \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{b}. \quad (\text{A.17})$$

A.4 Regularization

The HOA-encoded sound fields are obtained by means of Shape Matching (Eq. A.14). The matrix $(\mathbf{H}\mathbf{H}^H)$ can be ill-conditioned, especially at low frequencies, where the transfer functions between the microphones and loudspeakers can be virtually equivalent, making the matrix to be inverted $(\mathbf{H}\mathbf{H}^H)$ rank deficient. Inverting such a matrix results in large norm, highly unstable encoding matrices \mathbf{E} . Regularization is a tool intended to provide solutions which are reliable and stable. The most common regularization method is called Tikhonov, which comes from controlling the norm of the solution during the formulation of the optimization problem. Applying Tikhonov to Eq. A.14 results in

$$\mathbf{E} = \mathbf{C}\mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \lambda\mathbf{I}_Q)^{-1}, \quad (\text{A.18})$$

where λ is the regularization parameter and \mathbf{I}_Q is the identity matrix of size $[Q \times Q]$. The value of λ that we have used in this study to encode the sound fields is $\lambda = 0.4$.

Appendix B

Ethics



Australian Hearing Human Research Ethics Committee

APPROVAL FOR RESEARCH INVOLVING HUMAN SUBJECTS

APPROVAL NUMBER: AHHREC2017-3

Project Number	XR1.1.1c
Project Title	Evaluation of cochlear implant listening in realistic conditions
Classification	<i>Class 2: Project involving low risk.</i>
Principal Investigators authorized to conduct research	Javier Badajoz Davila, Jörg Buchholz
Date Approved	10/2/2017
Approval Method	Approved at the meeting of the full Committee.

This approval is based on the information contained in the ethics application that was presented to the Committee on 31/1/2017 and is conditional upon your continuing compliance with the National Statement on Ethical Conduct in Human Research (2007) available at:

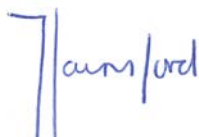
<https://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research> .

A duplicate set of the documents is enclosed for your records.

Annual reporting to the Committee on progress of the project is required including a final report when the work is completed or discontinued for any reason. Reminders will be sent when progress reports are due.

The Committee expects to be notified of any changes to the approved protocol or other issues that may have an impact on the ethics of the project either by means of the annual progress reports (checklists) or as an application for variation. Adverse or unforeseen events that affect the continued ethical acceptability of the project should be reported to the Chairman immediately.

All future correspondence relating to the ethical aspects of this project must quote the above Approval Number.



Dr Tim Gainsford
Operations & Finance Manager, NAL
and AHHREC Secretary

From: Ethics Secretariat [<mailto:ethics.secretariat@mq.edu.au>]
Sent: Wednesday, 19 July 2017 4:18 PM
To: Buchholz, Jorg
Subject: Re: Ethics Application for Externally Approved Project

Dear Dr Buchholz,

Re: Evaluation of cochlear implant listening in realistic conditions (Ref: 5201700509)

Thank you for your externally approved ethics application.

This project has received ethics approval from the Australian Hearing HREC and none of the procedures are being conducted at Macquarie University.

Please take this email as confirmation that the project has been noted by the Macquarie University Research Office.

Many thanks for providing this information for our records. No further action is required. Any amendments must be submitted to the approving HREC.

Please do not hesitate to contact the Ethics Secretariat if you have any concerns.



COCHLEAR
IMPLANT PROGRAM
An RIDBC service

361 – 365 North Rocks Road
North Rocks NSW 2151
T: 1300 658 981
E: scicadmin@scic.org.au
W: www.ridbc.org.au/scic

24 May 2019

Dr Jörg Buchholz
Auditory Scientist (CORE)
Macquarie University
Department of Linguistics
NSW 2113 Australia

Dear Dr Buchholz,

Re.: Your application to conduct research with the SCIC Cochlear Implant Program titled “Effect of reverberation on speech intelligibility in cochlear implant recipients considering realistic sound environments”

This letter is to confirm that the application named above was considered by the SCIC Cochlear Implant Program Research Governance Committee and approval had been granted in May 2017 as follows:

Approval was granted for conduct of the project in the SCIC services in accordance with the research protocol, associated participant information and consent forms, and proof of ethical approval by a properly constituted Human Research Ethics Committee, as submitted with your application.

Please note that this letter is primarily for documentation purposes.

Yours sincerely,

Dr. WaiKong Lai
Senior Research Engineer
(for the Research Committee)

05/10/2018

Dear Dr Buchholz,

Reference No: 5201835564572

Project ID: 3556

Title: Speech intelligibility performance with unilateral and bilateral cochlear implant users considering realistic sound environments

Thank you for submitting the above application for ethical review. The Human Sciences Subcommittee has considered your application.

I am pleased to advise that ethical approval has been granted for this project to be conducted by Dr Joerg Buchholz, and other personnel: Mr Javier Badajoz Davila.

Please be reminded to forward approval from the Sydney Cochlear Implant Centre (SCIC) once it has been obtained for our records.

This research meets the requirements set out in the National Statement on Ethical Conduct in Human Research 2007, (updated July 2018).

Standard Conditions of Approval:

1. Continuing compliance with the requirements of the National Statement, available from the following website:
<https://nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018>.
2. This approval is valid for five (5) years, subject to the submission of annual reports. Please submit your reports on the anniversary of the approval for this protocol. You will be sent an automatic reminder email one week from the due date to remind you of your reporting responsibilities.
3. All adverse events, including unforeseen events, which might affect the continued ethical acceptability of the project, must be reported to the subcommittee within 72 hours.
4. All proposed changes to the project and associated documents must be submitted to the subcommittee for review and approval before implementation. Changes can be made via the [Human Research Ethics Management System](#).

The HREC Terms of Reference and Standard Operating Procedures are available from the Research Services website:
<https://www.mq.edu.au/research/ethics-integrity-and-policies/ethics/human-ethics>.

It is the responsibility of the Chief Investigator to retain a copy of all documentation related to this project and to forward a copy of this approval letter to all personnel listed on the project.

Should you have any queries regarding your project, please contact the [Faculty Ethics Officer](#).

The Human Sciences Subcommittee wishes you every success in your research.

Yours sincerely,



Dr Naomi Sweller

Chair, Human Sciences Subcommittee



COCHLEAR
IMPLANT PROGRAM
An RIDBC service

361 – 365 North Rocks Road
North Rocks NSW 2151
T: 1300 658 981
E: scicadmin@scic.org.au
W: www.ridbc.org.au/scic

29 October 2018

Dr Jörg Buchholz
Auditory Scientist (CORE)
Macquarie University
Department of Linguistics
NSW 2113 Australia

Dear Dr Buchholz,

Re.: Your application to conduct research with the SCIC Cochlear Implant Program titled “Speech intelligibility performance with unilateral and bilateral cochlear implant users considering realistic sound environments”

I am pleased to advise that the application named above has been considered by the SCIC Cochlear Implant Program Research Governance Committee and approval has been granted as follows:

Approval is granted for conduct of the project in the SCIC services in accordance with the research protocol, associated participant information and consent forms, and proof of ethical approval by a properly constituted Human Research Ethics Committee, as submitted with your application.

Please note, in accordance with your application, you are obliged to keep this Research Committee informed of any changes to the approved protocol for the project and to provide annual reports on the study's progress.

We wish you every success with the conduct of the project.

Yours sincerely,

Prof. Greg Leigh, AO, PhD, FACE
Director, RIDBC Renwick Centre
(for the Research Committee)

Bibliography

- Ali, Hussnain et al. (2014). "Evaluation of adaptive dynamic range optimization in adverse listening conditions for cochlear implants". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.4893334](https://doi.org/10.1121/1.4893334).
- ANSI S3.35 (2004). "American National Standard: Method of Measurement of Performance Characteristics of Hearing Aids under Simulated Real-Ear Working Conditions". *American National Standards Institute, New York*.
- ANSI S3.5 (1997). "American National Standard: Methods for calculation of the speech intelligibility index". *New York*.
- Arweiler, Iris and Jörg M. Buchholz (2011). "The influence of spectral characteristics of early reflections on speech intelligibility". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.3609258](https://doi.org/10.1121/1.3609258).
- Beechey, Timothy (2019). "Communication difficulty and effort in conversation". PhD thesis. Macquarie University.
- Beechey, Timothy, Jörg M. Buchholz, and Gitte Keidser (2019). "Eliciting Naturalistic Conversations: A Method for Assessing Communication Ability, Subjective Experience, and the Impacts of Noise and Hearing Impairment". *Journal of Speech, Language, and Hearing Research*. ISSN: 1092-4388. DOI: [10.1044/2018_jslhr-h-18-0107](https://doi.org/10.1044/2018_jslhr-h-18-0107).
- Bench, John, Ase Kowal, and John Bamford (1979). "The bkb (bamford-kowal-bench) sentence lists for partially-hearing children". *British Journal of Audiology* 13.3, pp. 108–112. ISSN: 03005364. DOI: [10.3109/03005367909078884](https://doi.org/10.3109/03005367909078884).
- Bertet, S, J Daniel, and S Moreau (2006). "3D Sound Field Recording With Higher Order Ambisonics-Objective Measurements and Validation of Spherical Microphone". *Audio Engineering Society Convention 120*, pp. 1–24.
- Best, Virginia et al. (2015). "An examination of speech reception thresholds measured in a simulated reverberant cafeteria environment". *International Journal of Audiology*. ISSN: 17088186. DOI: [10.3109/14992027.2015.1028656](https://doi.org/10.3109/14992027.2015.1028656).
- Best, Virginia et al. (2016). "Development and preliminary evaluation of a new test of ongoing speech comprehension". *International Journal of Audiology*. ISSN: 17088186. DOI: [10.3109/14992027.2015.1055835](https://doi.org/10.3109/14992027.2015.1055835).
- Bistafa, Sylvio R. and John S. Bradley (2000). "Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.428268](https://doi.org/10.1121/1.428268).

- Bolt, R. H. (1949). "Theory of speech masking in reverberation". *J. Acoust. Soc. Am.* 21.6, pp. 577–580. ISSN: NA. DOI: [10.1121/1.1906551](https://doi.org/10.1121/1.1906551).
- Bradley, J. S. (1986). "Speech intelligibility studies in classrooms". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.393908](https://doi.org/10.1121/1.393908).
- Bradley, J. S. and S. R. Bistafa (2002). "Relating speech intelligibility to useful-to-detrimental sound ratios (L)". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.1481508](https://doi.org/10.1121/1.1481508).
- Bradley, John S (2002). "Optimising sound quality for classrooms". *XX Encontro da SOBRAC, II Simpósio Brasileiro de Metrologia em Acústica e Vibrações–SIBRAMA*, Rio de Janeiro.
- Bronkhorst, A. W. and R. Plomp (2005). "The effect of head-induced interaural time and level differences on speech intelligibility in noise". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.2024170](https://doi.org/10.1121/1.2024170).
- Bronkhorst, Adelbert W (2000). "The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions". *Acta Acust. United with Acust.* 86.January 2000, pp. 117–128. ISSN: 16101928. DOI: [10.1306/74D710F5-2B21-11D7-8648000102C1865D](https://doi.org/10.1306/74D710F5-2B21-11D7-8648000102C1865D).
- Bronkhorst, Adelbert W. and Tammo Houtgast (1999). "Auditory distance perception in rooms". *Nature* 397.6719, pp. 517–520. ISSN: 00280836. DOI: [10.1038/17374](https://doi.org/10.1038/17374).
- Buchholz, Jörg M. and Adam Weisser (2019). *Ambisonics Recordings of Typical Environments (ARTE) Database (Version 1.0.0)*. DOI: [10.5281/zenodo.2261632](https://doi.org/10.5281/zenodo.2261632).
- Chen, Fei, Oldooz Hazrati, and Philipos C. Loizou (2013). "Predicting the intelligibility of reverberant speech for cochlear implant listeners with a non-intrusive intelligibility measure". *Biomedical Signal Processing and Control*. ISSN: 17468094. DOI: [10.1016/j.bspc.2012.11.007](https://doi.org/10.1016/j.bspc.2012.11.007).
- Cherry, E. Colin (1953). "Some Experiments on the Recognition of Speech, with One and with Two Ears". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.1907229](https://doi.org/10.1121/1.1907229).
- Chu, WT and ACC Warnock (2002). "Detailed directivity of sound fields around human talkers". *Institute for Research in Construction, National Research Council Canada, Tech. Rep* December. DOI: <http://dx.doi.org/10.4224/20378930>.
- Compton-Conley, Cynthia L. et al. (2008). "Performance of Directional Microphones for Hearing Aids: Real-World versus Simulation". *Journal of the American Academy of Audiology*. ISSN: 10500545. DOI: [10.3766/jaaa.15.6.5](https://doi.org/10.3766/jaaa.15.6.5).
- Cooke, Martin (2006). "A glimpsing model of speech perception in noise". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.2166600](https://doi.org/10.1121/1.2166600).
- Cord, Mary et al. (2007). "Disparity between clinical assessment and real-world performance of hearing aids". *Hearing Review*.
- Cord, Mary T. et al. (2002). "Performance of directional microphone hearing aids in everyday life". *Journal of the American Academy of Audiology*. ISSN: 10500545.

- Cord, Mary T et al. (2004). "Relationship between laboratory measures of directional advantage and everyday success with directional microphone hearing aids." *Journal of the American Academy of Audiology*. ISSN: 1050-0545.
- Cox, Robyn M. and Genevieve C. Alexander (1995). "The abbreviated profile of hearing aid benefit". *Ear and Hearing*. ISSN: 15384667. DOI: [10.1097/00003446-199504000-00005](https://doi.org/10.1097/00003446-199504000-00005).
- Culling, John F. (2016). "Speech intelligibility in virtual restaurants". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.4964401](https://doi.org/10.1121/1.4964401).
- Daniel, Jérôme (2001). "Representation de champs acoustiques, application a la transmission et a la restitution de scenes sonores complexes dans un contexte multimedia (Acoustic field representation, application to the transmission and the reproduction of complex sound scenes in". PhD thesis, p. 319.
- Darwin, C J and R W Hukin (2000). "Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention." *The Journal of the Acoustical Society of America* 108.1, pp. 335–42. ISSN: 0001-4966. DOI: [10.1121/1.429468](https://doi.org/10.1121/1.429468).
- Dawson, Pam W, Stefan J Mauger, and Adam a Hersbach (2011). "Clinical evaluation of signal-to-noise ratio-based noise reduction in Nucleus® cochlear implant recipients." *Ear and hearing* 32.3, pp. 382–390. ISSN: 0196-0202. DOI: [10.1097/AUD.0b013e318201c200](https://doi.org/10.1097/AUD.0b013e318201c200).
- Desmond, Jill M., Leslie M. Collins, and Chandra S. Throckmorton (2014). "The effects of reverberant self- and overlap-masking on speech recognition in cochlear implant listeners". *The Journal of the Acoustical Society of America* 135.6, EL304–EL310. ISSN: 0001-4966. DOI: [10.1121/1.4879673](https://doi.org/10.1121/1.4879673).
- Falk, Tiago H., Chenxi Zheng, and Wai Yip Chan (2010). "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech". *IEEE Transactions on Audio, Speech and Language Processing*. ISSN: 15587916. DOI: [10.1109/TASL.2010.2052247](https://doi.org/10.1109/TASL.2010.2052247).
- Falk, Tiago H. et al. (2015). "Objective quality and intelligibility prediction for users of assistive listening devices: Advantages and limitations of existing tools". *IEEE Signal Processing Magazine*. ISSN: 10535888. DOI: [10.1109/MSP.2014.2358871](https://doi.org/10.1109/MSP.2014.2358871).
- Favrot, S. and J. M. Buchholz (2010). "LoRA: A loudspeaker-based room auralization system". *Acta Acustica united with Acustica*. ISSN: 16101928. DOI: [10.3813/AAA.918285](https://doi.org/10.3813/AAA.918285).
- Favrot, Sylvain et al. (2011). "Mixed-order Ambisonics recording and playback for improving horizontal directionality". *131st Audio Engineering Society Convention 2011, October 20, 2011 - October 23*. Vol. 2, pp. 641–647.
- Festen, Joost M. and Reinier Plomp (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. ISSN: 0001-4966. DOI: [10.1121/1.400247](https://doi.org/10.1121/1.400247).

- Fu, Qian Jie and Geraldine Nogaki (2005). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing". *JARO - Journal of the Association for Research in Otolaryngology* 6.1, pp. 19–27. ISSN: 15253961. DOI: [10.1007/s10162-004-5024-3](https://doi.org/10.1007/s10162-004-5024-3).
- Galster, Jason Alan (2007). "The Effect of Room Volume on Speech Recognition in Enclosures with Similar Mean Reverberation Time". PhD thesis. Vanderbilt University, Nashville, Tennessee.
- Galvez, Gino et al. (2012). "Feasibility of ecological momentary assessment of hearing difficulties encountered by hearing aid users". *Ear and Hearing*. ISSN: 01960202. DOI: [10.1097/AUD.0b013e3182498c41](https://doi.org/10.1097/AUD.0b013e3182498c41).
- Gatehouse, Stuart and Iliam Noble (2004). "The Speech, Spatial and Qualities of Hearing Scale (SSQ)". *International Journal of Audiology*. ISSN: 14992027. DOI: [10.1080/14992020400050014](https://doi.org/10.1080/14992020400050014).
- Gelfand, Stanley and Shlomo Silman (1979). "Effects of small room reverberation upon the recognition of some consonant features". *Journal of the Acoustical Society of America* 66.1, pp. 22–29. ISSN: NA. DOI: [10.1121/1.383075](https://doi.org/10.1121/1.383075).
- Gifford, René H. and Lawrence J. Revit (2011). "Speech Perception for Adult Cochlear Implant Recipients in a Realistic Background Noise: Effectiveness of Preprocessing Strategies and External Options for Improving Speech Recognition in Noise". *Journal of the American Academy of Audiology*. ISSN: 10500545. DOI: [10.3766/jaaa.21.7.3](https://doi.org/10.3766/jaaa.21.7.3).
- Goldsworthy, Ray L. and Julie E. Greenberg (2004). "Analysis of speech-based speech transmission index methods with implications for nonlinear operations". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.1804628](https://doi.org/10.1121/1.1804628).
- Goorevich, M and M Batty (2005). "A new real-time research platform for the Nucleus® 24 and Nucleus® Freedom™ cochlear implants". *Conference on Implantable Auditory Prostheses (CIAP)*.
- Greenberg, Steven et al. (2004). *Speech processing in the auditory system*. Vol. 18. Springer. ISBN: 0387005900.
- Grimm, Giso, Stephan Ewert, and Volker Hohmann (2015). "Evaluation of spatial audio reproduction schemes for application in hearing aid research". *Acta Acustica united with Acustica*. ISSN: 16101928. DOI: [10.3813/AAA.918878](https://doi.org/10.3813/AAA.918878).
- Hagerman, B. (1982). "Sentences for Testing Speech Intelligibility in Noise". *Scandinavian Audiology*. ISSN: 01050397. DOI: [10.3109/01050398209076203](https://doi.org/10.3109/01050398209076203).
- Hartmann, W.M. (1983). "Localization of sound in rooms". *Jasa* 74.November, pp. 1380–1391. ISSN: 00014966. DOI: [10.1121/1.390163](https://doi.org/10.1121/1.390163).
- Hazrati, Oldooz, Jaewook Lee, and Philipos C. Loizou (2013). "Blind binary masking for reverberation suppression in cochlear implants". *The Journal of the Acoustical Society of America* 133.3, pp. 1607–1614. ISSN: 0001-4966. DOI: [10.1121/1.4789891](https://doi.org/10.1121/1.4789891).

- Hazrati, Oldooz and Philipos C. Loizou (2012). "The combined effects of reverberation and noise on speech intelligibility by cochlear implant listeners". *International Journal of Audiology* 51.January, pp. 437–443. ISSN: 1499-2027. DOI: [10.3109/14992027.2012.658972](https://doi.org/10.3109/14992027.2012.658972).
- Hazrati, Oldooz and Philipos C Loizou (2013). "Reverberation suppression in cochlear implants using a blind channel-selection strategy". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.4804313](https://doi.org/10.1121/1.4804313).
- Helms Tillery, Kate, Christopher A. Brown, and Sid P. Bacon (2012). "Comparing the effects of reverberation and of noise on speech recognition in simulated electric-acoustic listening". *The Journal of the Acoustical Society of America* 131.1, pp. 416–423. ISSN: 0001-4966. DOI: [10.1121/1.3664101](https://doi.org/10.1121/1.3664101).
- Hersbach, Adam A. et al. (2015). "Perceptual effect of reverberation on multi-microphone noise reduction for cochlear implants". *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. ISBN: 9781467369978. DOI: [10.1109/ICASSP.2015.7179094](https://doi.org/10.1109/ICASSP.2015.7179094).
- Hohmann, V. (2002). "Frequency analysis and synthesis using a Gammatone filter-bank". *Acta Acustica united with Acustica*. ISSN: 14367947.
- Hopkins, Kathryn and Brian C. J. Moore (2009). "The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise". *The Journal of the Acoustical Society of America* 125.1, pp. 442–446. ISSN: 0001-4966. DOI: [10.1121/1.3037233](https://doi.org/10.1121/1.3037233).
- Houtgast, T, H J M Steeneken, and R Plomp (1980). "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I. General Room Acoustics". *Acustica* 46.1, pp. 60–72.
- Hu, Yi and Kostas Kokkinakis (2014). "Effects of early and late reflections on intelligibility of reverberated speech by cochlear implant listeners." *The Journal of the Acoustical Society of America* 135.1, EL22–8. ISSN: 1520-8524. DOI: [10.1121/1.4834455](https://doi.org/10.1121/1.4834455).
- IEC (2003). 60268–16-2003 *Sound system equipment — Part 16: Objective rating of speech intelligibility by speech transmission index*. ISBN: 978 0 580 76163 8.
- ISO 3382-1 (2009). *Acoustics - Measurement of room acoustic parameters. Part 1: Performance spaces*. DOI: [10.1017/CB09781107415324.004](https://doi.org/10.1017/CB09781107415324.004).
- Jacobsen, Finn et al. (2013). *Fundamentals of Acoustics - Note no 31200*. Tech. rep. Technical University of Denmark.
- Jerger, J (2009). "Ecologically valid measures of hearing aid performance". *Starkey Audiology Series* 1.1, pp. 1–4.
- Jørgensen, Søren and Torsten Dau (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*. ISSN: 0001-4966. DOI: [10.1121/1.3621502](https://doi.org/10.1121/1.3621502).
- Kates, J M (2008). *Digital Hearing Aids*. Plural Pub. ISBN: 9781597563178. URL: <https://books.google.com.au/books?id=Pu07IQAACAAJ>.

- Keidser, Gitte et al. (2013). "An algorithm that administers adaptive speech-in-noise testing to a specified reliability at selectable points on the psychometric function". *International Journal of Audiology* 52.11, pp. 795–800. ISSN: 14992027. DOI: [10.3109/14992027.2013.817688](https://doi.org/10.3109/14992027.2013.817688).
- Kelly M. Miles et al. (2019). "Development of an intelligibility test based on conversational speech produced in realistic noise". Submitted.
- Kiessling, J. et al. (2003). "Candidature for and delivery of audiological services: special needs of older people". *International Journal of Audiology*. ISSN: 1499-2027. DOI: [10.3109/14992020309074650](https://doi.org/10.3109/14992020309074650).
- Killion, Mead et al. (1998). "Real-world performance of an lie directional microphone". *Hearing Journal* 51, pp. 24–39.
- Kokkinakis, Kostas, Oldoos Hazrati, and Philipos C Loizou (2011). "A channel-selection criterion for suppressing reverberation in cochlear implants." *The Journal of the Acoustical Society of America* 129.5, pp. 3221–32. ISSN: 1520-8524. DOI: [10.1121/1.3559683](https://doi.org/10.1121/1.3559683).
- Kokkinakis, Kostas and Philipos C. Loizou (2011). "The impact of reverberant self-masking and overlap-masking effects on speech intelligibility by cochlear implant listeners (L)". *The Journal of the Acoustical Society of America* 130.3, pp. 1099–1102. ISSN: 0001-4966. DOI: [10.1121/1.3614539](https://doi.org/10.1121/1.3614539). URL: <http://asa.scitation.org/doi/10.1121/1.3614539>.
- Kressner, Abigail Anne, Adam Westermann, and Jörg M. Buchholz (2018). "The impact of reverberation on speech intelligibility in cochlear implant recipients". *The Journal of the Acoustical Society of America* 144.2, pp. 1113–1122. ISSN: 0001-4966. DOI: [10.1121/1.5051640](https://doi.org/10.1121/1.5051640).
- Lavandier, Mathieu and John F. Culling (2008). "Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.2871943](https://doi.org/10.1121/1.2871943).
- Lombard, E (1911). "Le signe de l'élévation de la voix (translated from French)". *Annales des maladies de l'oreille et du larynx*.
- Lu, Youyi and Martin Cooke (2008). "Speech production modifications produced by competing talkers, babble, and stationary noise". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.2990705](https://doi.org/10.1121/1.2990705).
- Mauger, Stefan J. et al. (2014). "Clinical evaluation of the Nucleus®6 cochlear implant system: Performance improvements with SmartSound iQ". *International Journal of Audiology* 53.8, pp. 564–576. ISSN: 17088186. DOI: [10.3109/14992027.2014.895431](https://doi.org/10.3109/14992027.2014.895431).
- Mueller, Martin F. et al. (2012). "Localization of virtual sound sources with bilateral hearing aids in realistic acoustical scenes". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.4705292](https://doi.org/10.1121/1.4705292).
- Nabelek, Anna K and James M Pickett (1974). "Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and

- hearing-impaired listeners". *Journal of Speech, Language, and Hearing Research* 17.4, pp. 724–739. ISSN: 1092-4388.
- Nábélek, Anna K. and Larry Robinette (1978). "Reverberation as a parameter in clinical testing". *International Journal of Audiology* 17.3, pp. 239–259. ISSN: 14992027. DOI: [10.1080/00206097809086955](https://doi.org/10.1080/00206097809086955).
- Nelson, Peggy B. et al. (2003). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners". *The Journal of the Acoustical Society of America* 113.2, pp. 961–968. ISSN: 0001-4966. DOI: [10.1121/1.1531983](https://doi.org/10.1121/1.1531983).
- Nijs, Lau and Monika Rychtáriková (2011). "Calculating the optimum reverberation time and absorption coefficient for good speech intelligibility in classroom design using U50". *Acta Acustica united with Acustica*. ISSN: 16101928. DOI: [10.3813/AAA.918390](https://doi.org/10.3813/AAA.918390).
- Oreinos, C (2015a). "Virtual acoustic environments for the evaluation of hearing devices". PhD thesis. PhD thesis. Macquarie University, Sydney, Australia. DOI: <http://hdl.handle.net/1959.14/1269481>.
- Oreinos, Chris (2015b). "Objective analysis of ambisonics for hearing aid applications: Effect of listener's head, room reverberation, and directional microphones". *The Journal of the Acoustical Society of America* 137.6, pp. 3447–3465. DOI: [10.1121/1.4919330](https://doi.org/10.1121/1.4919330).
- Oreinos, Chris and Jörg M. Buchholz (2016). "Evaluation of Loudspeaker-Based Virtual Sound Environments for Testing Directional Hearing Aids". *Journal of the American Academy of Audiology*. ISSN: 10500545. DOI: [10.3766/jaaa.15094](https://doi.org/10.3766/jaaa.15094).
- Pearsons, K.S, R.L Bennett, and S. Fidell (1976). "Speech levels in various environments". *Bolt Beranek and Newman*.
- Peissig, Jürgen and Birger Kollmeier (2002). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.418150](https://doi.org/10.1121/1.418150).
- Pichora-Fuller, M. Kathleen and Gurjit Singh (2006). "Effects of Age on Auditory and Cognitive Processing: Implications for Hearing Aid Fitting and Audiologic Rehabilitation". *Trends in Amplification*. ISSN: 19405588. DOI: [10.1177/108471380601000103](https://doi.org/10.1177/108471380601000103).
- Pinheiro, Jose C. and Douglas M. Bates (2000). *Mixed-Effects Models in S and S-Plus*. ISBN: 0-387-98957-5.x. DOI: [10.1007/b98882](https://doi.org/10.1007/b98882).
- Poissant, Sarah F, Nathaniel a Whitmal, and Richard L Freyman (2006). "Effects of reverberation and masking on speech intelligibility in cochlear implant simulations." *The Journal of the Acoustical Society of America* 119.3, pp. 1606–1615. ISSN: 00014966. DOI: [10.1121/1.2168428](https://doi.org/10.1121/1.2168428).
- Qin, Michael K. and Andrew J. Oxenham (2003). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers". *The Journal of the Acoustical Society of America* 114.1, pp. 446–454. ISSN: 0001-4966. DOI: [10.1121/1.1579009](https://doi.org/10.1121/1.1579009).

- Rana, Baljeet et al. (2017). "Bilateral Versus Unilateral Cochlear Implantation in Adult Listeners: Speech-On-Speech Masking and Multitalker Localization". *Trends in Hearing*. ISSN: 23312165. DOI: [10.1177/2331216517722106](https://doi.org/10.1177/2331216517722106).
- Rychtáriková, Monika et al. (2009). "Binaural sound source localization in real and virtual rooms". *AES: Journal of the Audio Engineering Society*. ISSN: 15494950.
- Sabine, W C (1993). *Collected Papers on acoustics (Originally 1921)*.
- Santos, João F. and Tiago H. Falk (2014). "Updating the SRMR-CI metric for improved intelligibility prediction for cochlear implant users". *IEEE/ACM Transactions on Audio Speech and Language Processing*. ISSN: 23299290. DOI: [10.1109/TASLP.2014.2363788](https://doi.org/10.1109/TASLP.2014.2363788).
- Santos, João F. et al. (2013). "Objective speech intelligibility measurement for cochlear implant users in complex listening environments". *Speech Communication*. ISSN: 01676393. DOI: [10.1016/j.specom.2013.04.001](https://doi.org/10.1016/j.specom.2013.04.001).
- Schröder, M (1954). "Eigenfrequenzstatistik und Anregungsstatistik in Räumen". *Acta Acustica united with Acustica* 4.
- Seeber, Bernhard U., Stefan Kerber, and Ervin R. Hafter (2010). "A system to simulate and reproduce audio-visual environments for spatial hearing research". *Hearing Research*. ISSN: 03785955. DOI: [10.1016/j.heares.2009.11.004](https://doi.org/10.1016/j.heares.2009.11.004).
- Smeds, Karolina, Florian Wolters, and Martin Rung (2015). "Estimation of Signal-to-Noise Ratios in Realistic Sound Scenarios". *Journal of the American Academy of Audiology*. ISSN: 10500545. DOI: [10.3766/jaaa.26.2.7](https://doi.org/10.3766/jaaa.26.2.7).
- Taal, Cees H. et al. (2011). "An algorithm for intelligibility prediction of time-frequency weighted noisy speech". *IEEE Transactions on Audio, Speech and Language Processing*. ISSN: 15587916. DOI: [10.1109/TASL.2011.2114881](https://doi.org/10.1109/TASL.2011.2114881).
- Van Hoesel, Richard (2011). "Auditory prostheses: New horizons". Ed. by Richard R. Zeng, Fan-Gang, Popper, Arthur N., Fay. Vol. 39. Springer Science & Business Media. Chap. 2.
- Van Hoesel, Richard and Richard S. Tyler (2003). "Speech perception, localization, and lateralization with bilateral cochlear implants". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.1539520](https://doi.org/10.1121/1.1539520).
- Walden, B E et al. (2000). "Comparison of benefits provided by different hearing aid technologies". *Journal of the American Academy of Audiology*.
- Warton, David I. and Francis K C Hui (2011). "The arcsine is asinine: The analysis of proportions in ecology". *Ecology*. ISSN: 00129658. DOI: [10.1890/10-0340.1](https://doi.org/10.1890/10-0340.1).
- Watkins, Greg D., Brett A. Swanson, and Gregg J. Suaning (2018). "An evaluation of output signal to noise ratio as a predictor of cochlear implant speech intelligibility". *Ear and Hearing*. ISSN: 15384667. DOI: [10.1097/AUD.0000000000000556](https://doi.org/10.1097/AUD.0000000000000556).
- Weisser, Adam and Jörg M. Buchholz (2019). "Conversational speech levels and signal-to-noise ratios in realistic acoustic conditions". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.5087567](https://doi.org/10.1121/1.5087567).
- Weisser, Adam et al. (2019). "The Ambisonic Recordings of Typical Environments (ARTE) Database". *Acta Acustica united with Acustica*. DOI: [10.3813/aaa.919349](https://doi.org/10.3813/aaa.919349).

- Westermann, Adam and Jörg M. Buchholz (2015). "The effect of spatial separation in distance on the intelligibility of speech in rooms". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.4906581](https://doi.org/10.1121/1.4906581).
- Whitmal, Nathaniel A. and Sarah F. Poissant (2009). "Effects of source-to-listener distance and masking on perception of cochlear implant processed speech in reverberant rooms". *The Journal of the Acoustical Society of America* 126.5, pp. 2556–2569. ISSN: 0001-4966. DOI: [10.1121/1.3216912](https://doi.org/10.1121/1.3216912).
- Wilson, Blake S. and Michael F. Dorman (2007). "The surprising performance of present-day cochlear implants". *IEEE Transactions on Biomedical Engineering*. ISSN: 00189294. DOI: [10.1109/TBME.2007.893505](https://doi.org/10.1109/TBME.2007.893505).
- Wolfe, Jace et al. (2015). "Benefits of Adaptive Signal Processing in a Commercially Available Cochlear Implant Sound Processor". *Otology and Neurotology*. ISSN: 15374505. DOI: [10.1097/MAO.0000000000000781](https://doi.org/10.1097/MAO.0000000000000781).
- Wu, Yu Hsiang et al. (2018). "Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss". *Ear and Hearing*. DOI: [10.1097/AUD.0000000000000486](https://doi.org/10.1097/AUD.0000000000000486).
- Xia, Jing et al. (2018). "Effects of reverberation and noise on speech intelligibility in normal-hearing and aided hearing-impaired listeners". *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: [10.1121/1.5026788](https://doi.org/10.1121/1.5026788).
- Zeng, Fan Gang et al. (2008). "Cochlear Implants: System Design, Integration, and Evaluation". *IEEE Reviews in Biomedical Engineering*. ISSN: 19411189. DOI: [10.1109/RBME.2008.2008250](https://doi.org/10.1109/RBME.2008.2008250).