Modelling the effect of hearing impairment for binaural speech intelligibility in noise

Thibault Vicente

N°d'ordre NNT : 2021LYSET004

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON
opéré au sein de

L'École Nationale des Travaux Public de l'État
En cotutelle internaltionale avec

Macquarie Universtity
Faculty of Medecine, Health and Human Sciences
Department of Linguitics

École Doctorale N°162
Mécanique, Énergétique, Génie civil, Acoustique
Spécialité / discipline de doctorat : Acoustique, Psychoacoustique

Soutenue publiquement le 22/06/2021, par :

Thibault Vicente

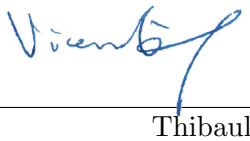
Modelling the effect of hearing impairment for binaural speech intelligibility in noise
Modélisation de l'effet de la perte auditive sur la compréhension de la parole dans le bruit en écoute binaurale

Devant le jury composé de :

Lorenzi, Christian, Pr	Ecole Normale Supérieure, Paris	Président
Akeroyd, Michael, Pr.	University of Notthingham, UK	Rapporteur
Brand, Thomas, Ass. Pr.	Carl von Ossietzky Universität Oldenburg, DE	Rapporteur
Colburn, Steven, Pr. Emer.	Boston University, USA	Rapporteur
Fels, Janina, Pr.	RWTH Aachen University, DE	Examinatrice
Lavandier, Mathieu CR1 HDR	ENTPE, Lyon	Directeur de thèse
Buchholz, Jörg Ass. Pr.	Macquarie University, AU	Directeur de thèse

Typeset in LaTeX 2_ε.

This work has not previously been submitted for a degree or diploma in any university. This thesis is being submitted to Macquarie University and the University of Lyon in accordance with the Cotutelle agreement dated 16/11/2018. 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. Ethical clearance was received from the Macquarie University Human Sciences Ethics Committee. The letter of acceptance with the approval number 5201955629923 can be found in Appendix D



03/08/2021

Thibault Vicente

Declaration of contributions to publications

Chapter III presents a modelling study including a data collection that was conducted by myself under the supervision of Mathieu Lavandier. The original matlab code of the model was written by Mathieu Lavandier, which was then adapted by myself for the study. Some of the data used for the sensitivity analysis were shared by John Culling and Stephan Ewert. I adapted a matlab code written by a previous student of Mathieu Lavandier to create the stimuli of the experiment. I was the experimenter during the data collection. The data analysis was done by myself after Mathieu Lavandier taught me how to use the statistical software (STATISTICA). The sensitivity analysis was implemented and interpreted by myself. I wrote the first draft of the published paper then I changed it according to the input of Mathieu Lavandier before the submission to Hearing Research. I carried out the publication process with inputs from Mathieu Lavandier.

Chapter IV shows a modelling study that was done by myself under the supervision of Mathieu Lavandier and Jörg Buchholz. The original model was coded by Mathieu Lavandier, but the changes presented in the chapter were done by myself. The internal noise formula presented in the chapter was obtained after long dicussions between the three of us using the outcomes of my literature review on the internal noise. The data used for the chapter were shared by Jörg Buchholz and Baljeet Rana. I wrote the first draft of the published paper and then changed it considering the comments and suggestions of Mathieu Lavandier and Jörg Buchholz before submitting the manuscript to the Journal of the Acoustical Society of America. I carried out the publication process with inputs from Jörg Buchholz and Mathieu Lavandier.

Chapter V presents an experimental and modelling study that was designed by myself under the supervision of Jörg Buchholz and Mathieu Lavandier. Regarding the experimental study, I designed the tested conditions with inputs from my supervisors. Kelly Miles and Jörg Buchholz helped me to get the ethic clearance for the data collection. The TFS-AF test (measuring interaural time difference sensitivity) used in the study was developed by Christian Füllgrabe and colleagues. The software to measure intelligibility thresholds was developed by Jörg Buchholz and slighlty changed by myself for our study. The original monaural stimuli were created by Jörg Buchholz in a previous study. The impulse responses were from a previous study by Mathieu Lavandier. I generated the binaural stimuli with this material. I carried out the recruitment of the participants with the help of Kelly Miles, Julie Beadle, Jörg Buchholz and Katrina Freeston. Katrina Freeston was the experimenter because we needed an Australian audiologist to test native Australian-English normal-hearing and hearing-impaired participants. I carried out the data analysis with the input of Jaime Undurraga for the statistics. I handled the modelling study considering the recommendations of my supervisors. I wrote the first draft of the submitted paper and changed it according to the comments of Mathieu Lavandier and Jörg Buchholz before submitting the manuscript to the Journal of the Acoustical Society of America. I am carrying out the publication process of the journal (and will receive inputs from my supervisors).

The authorship contributions forms can be found at the end of the manuscript in Appendix E.

Acknowledgements

First of all, I would like to sincerely thank the examiners of my manuscript for taking their time to review my research work. I would like to acknowledge the anonymous reviewers for their helpful comments on two of my chapters during the journal publication processes. I am grateful to the researchers who helped me during my PhD without asking anything in return: John Culling and Stephan Ewert for sharing their data; Emmanuel Gourdon for his valued advice on the sensitivity analysis; Kelly Miles, Julie Beadle and the MARCS Institute for their help in the data collection; Christian Füllgrabe and Brian Moore for sharing the TFS-AF test. I would like to acknowledge the project fundings from Fondation pour l'Audition (Speech2Ears grant), Macquarie University (through an international research excellence scholarship, iMQRES) and the LabEx CeLyA (ANR-10-LABX-0060/ANR-16-IDEX-0005).

I would like to deeply thank my supervisors, Jörg and Mat, for their help during the project but also for all the things we shared that were definitely not related to research. I could not have imagined a better supervision, they are complementing each other that makes them a amazing couple of supervisors. Lucky the next PhD student that will be supervised by these two chaps.

I have also a thought for:

- my family who has always supported me in my choice, especially my parents although they were half happy when I told them that I was going to live in Australia for one year.
- my friends because they all are an amazing support to me. I will not try to list them as a lot of people do because I think this is one of the first reasons for friend divorce.
- my flatemates I lived with during my PhD: Gab, Tangi, Tutu, Antoine, Mathis, Ben, Christie and this bad-ass half-dog half-cat called Wriggles. Sharing an accommodation is not always easy but we always managed to make it as simple as we can and this was so cool.
- the party animals crew that I met in Australia. Monday dinners and parties were good with you !
- Theo and Zoe who prefer to stay in Sydney at least until the end of the pandemic, you likely have made the good choice. My year in Australia would not have been as good without you.
- Gab, Nagromz, Nomiz and Le Glé who visited me when I was in Australia. This was epic and I am still wondering if Nomiz has recovered of this journey.
- Ben and Christie because they bring me everywhere in Sydney to show me the best place in town. I learnt a lot thanks to you I am grateful for this. I am stoked I met you and I do not forget our promise: we will see each other.
- Matthieu and Pauline for the Friday nights and more.
- Thibaud Leclère for having a look to my introducing chapters.
- Ilias because you pushed me to find a job in London and I am glad to join you !
- My colleagues from the French and Australian labs, because it was always a pleasure to come working at the office.

List of Publications

Peer-reviewed publications

- Vicente T., Buchholz J. M., Lavandier M., *The contribution of binaural unmasking to speech intelligibility in noise for normal−hearing and hearing−impaired listeners.* submitted (04/03/2020) to The Journal of the Acoustical Society of America
- Vicente T., Lavandier M., Buchholz J. M. *A binaural model implementing an internal noise to predict the effect of hearing impairment on speech intelligibility in non−stationary noises.* The Journal of the Acoustical Society of America **145** (5), 3305-3317 (2020)
- Vicente T., Lavandier M., *Further validation of a binaural model predicting speech intelligibility against envelope−modulated noise interferers.* Hearing research **390**, 107937 (2020)

Conference presentations during the cotutelle

- Vicente T., Buchholz J. M., Lavandier M., *Evaluation of the effect of interaural time difference sensitivity on speech intelligibility in noise for normal-hearing and hearing-impaired listeners.* e-Forum Acusticum, Lyon, FR, 7−11 December 2020
- Vicente T., Lavandier M., Buchholz J. M. *A model predicting the effect of hearing impairment and noise level on speech intelligibility.* 23$^{\text{rd}}$ International Congress of Acoustics, Aachen, DE, 9−13 September 2019
- Vicente T., Lavandier M., Buchholz J. M. *A model predicting the effect of hearing impairment on binaural speech intelligibility in noise.* 11$^{\text{th}}$ AABBA Meeting, Vienna, AT, 19−20 February 2019
- Vicente T., Lavandier M., Buchholz J. M. *A model predicting the effect of hearing impairment on binaural speech intelligibility in noise.* 11$^{\text{th}}$ Speech in Noise Workshop, Ghent, BE, 10−11 January 2019

Conference presentations before the cotutelle

- Vicente T., Lavandier M., Buchholz J. M., *Développement d'un modèle binaural de prédiction de l'intelligibilité de la parole dans le bruit pour des auditeurs malentendants.* 11$^{\text{th}}$ Days of the Young Research in Hearing, Musical acoustics and Signal processing, Brest, FR, 6−8 June 2018
- Vicente T., Lavandier M., Buchholz J. M., *Développement d'un modèle binaural de prédiction de l'intelligibilité de la parole dans le bruit pour des auditeurs malentendants.* 14$^{\text{th}}$ French Congress of Acoustics, Le Havre, FR, 23−27 April 2018
- Vicente T., Lavandier M., *Révision d'un modèle binaural prédisant l'intelligibilité de la parole en présence de sources de bruit non−stationnaire.* 14$^{\text{th}}$ French Congress of Acoustics, Le Havre, FR, 23−27 April 2018
- Vicente T., Lavandier M., *Revision of a binaural model predicting speech intelligibility against envelope−modulated noise.* 10$^{\text{th}}$ Speech in Noise Workshop, Glasgow, UK, 11−12 January 2018

Lists of acronyms, notations and symbols

Acronyms

4FAHL	Four-frequency average hearing loss
BMLD	Binaural masking level difference
BRIR	Binaural room impulse response
EC	Equalization-cancellation
ERB	Equivalent rectangular bandwidth
ERD	Equivalent rectangular duration
HI	Hearing-impaired
HRIR	Head-related impulse response
IHC	Inner hair cell
ILD	Interaural level difference
IMBM	Ideal monaural better-ear mask
IPD	Interaural phase difference
ITD	Interaural time difference
NAL-RP	National Acoustics Laboratories - Revised Profound
NH	Normal-hearing
OHC	Outer hair cell
dB SL	dB sensation level
SEIR	Spectral-envelope impulse response
SII	Speech intelligibility index
SNR	Signal-to-noise ratio
SRM	Speech release from masking
SRT	Speech reception threshold
SSN	Speech-shaped noise
STI	Speech transmission index
TFS-AF test	Temporal fine structure adaptive frequency test
VS	Noise-vocoded speech masker

Notations

BE	Duration of the Hann window used for computing the better-ear SNR in the model (specific to Chapter III)
BU	Duration of the Hann window used for computing the binaural unmasking advantage in the model (specific to Chapter III)

Ceiling..... Maximum better-ear SNR allowed by frequency band and time frame in the model (specific to Chapter III)

IPD_{Freq} ..... Upper frequency to detect an IPD (specific to Chapter V)

$MeanErr$.... Mean absolute error between data and predictions

$MaxErr$..... Maximal absolute error between data and predictions

$RMSErr$.... Root-mean-square error between data and predictions

SpecSamp.... Model spectral sampling (specific to Chapter III)

Symbols

N..... External noise level (used only in formula)

N_{int} ..... Internal noise level (used only in formula)

r..... Pearson's correlation coefficient between data and predictions

r_s..... Spearman's correlation coefficient between data and predictions

T..... Pure-tone audiogram (used only in formula)

α..... Absorption coefficient of the room (used only in formula)

Γ..... Transformation to convert the hearing loss from dB HL to dB SPL (used only in formula)

η..... Proportion of outer hair cell loss

ρ..... Noise interaural coherence (used only in formula)

σ_δ ..... Time jitter

σ_ϵ..... Level jitter

θ..... interaural phase difference (used only in formula)

Abstract

Understanding speech amongst others is a challenging and daily situation occurring in public transport, food court, pub... A listener is able to segregate the target speech from the competing sources, so-called maskers, when they are spatially separated thanks to auditory mechanisms that use interaural level and time differences (ILD and ITD) of the signals. These mechanisms are well known as better-ear listening and binaural unmasking, respectively. However, their benefits are degraded when a listener suffers from hearing loss so that speech intelligibility can be greatly reduced. The motivation for this PhD is to develop a binaural speech intelligibility model that can account for the effects related to hearing loss in complex, realistic scenarios, including reverberation and competing masking sources. This model would help to better understand what aspects of hearing loss impact speech intelligibility.

The project has been structured in three main studies. The first study was about the optimization of a binaural speech intelligibility model for normal-hearing (NH) listeners applied to datasets testing auditory better-ear listening and binaural unmasking in isolation and combination, thus, hearing impairment was discarded in a first place. The model inputs are the speech and masker signals alone. They are decomposed per time frame and frequency band. Then, better-ear listening is modelled by taking the higher signal-to-noise ratio (SNR) between ears and the binaural unmasking advantage is estimated using a formula previously developed in the literature. The values are integrated across frequency, averaged across time and summed to provide a binaural ratio that can be compared to speech intelligibility threshold using a scaling method. It must be seen as a deep introduction of the original model that was further developed in the following studies to take into account hearing loss.

The second study investigated the influence of hearing loss on better-ear listening. The main outcome was the design of an internal noise level that can account for the effect of individual reduced audibility. The internal noise level is spectrally shaped on the listener's audiogram and relies on the external stimulus level. In this version, the better-ear SNR is computed using the higher level between internal noise and external masker and the binaural unmasking advantage is computed only if the external signal levels are above the internal noise level at each ear. The model implementation was tested using three experiments from the literature involving NH and hearing-impaired (HI) listeners and varying the spatial separation of the sources, the masker type as well as the sensation level.

The last study highlights the contribution of binaural unmasking to speech intelligibility for NH and HI listeners. For this purpose, a dataset was collected during the project including ITD sensitivity and speech intelligibility measurements. Speech intelligibility was measured varying the presentation level of the signals, the masker type, the difference in ITD between speech and masker as well as the reverberation of the room. A new model version was developed making the two jitters of the formula computing the binaural unmasking advantage in the model dependent of the external stimulus level. This allowed to account for the effect of low presentation level on binaural unmasking for NH listeners.

Contents

Declaration v

Acknowledgements vii

List of Publications ix

Lists of acronyms, notations and symbols xi

Abstract xiii

List of Figures xvii

List of Tables xxi

I Introduction 1

II Background 3
II.1 Speech intelligibility in noise . . . 3
II.1.1 Signal-to-noise ratio and intelligibility . . . 3
II.1.2 Dip listening . . . 4
II.1.3 Spatial release from masking . . . 4
II.1.4 Effect of room reverberation . . . 8
II.1.5 Effect of hearing impairment . . . 9
II.2 Speech intelligibility models . . . 13
II.2.1 Monaural models for hearing-impaired listeners or hearing-aid processing 13
II.2.2 Binaural models tested only with normal-hearing listeners . . . 18
II.2.3 Binaural models for hearing-impaired listeners or hearing-aid processing . 19
II.2.4 Summary and outlook . . . 23
II.3 The role of this PhD project . . . 24

III Further validation of a binaural model predicting speech intelligibility against envelope-modulated noises 27
III.1 Introduction . . . 27
III.2 Data . . . 28
III.2.1 Data sets used to test the model parameter . . . 28
III.2.2 Data set used to validate the revised model . . . 29
III.3 Model description . . . 30
III.4 Revision of the model . . . 31
III.4.1 A method inspired by a sensitivity analysis . . . 31
III.4.2 Results . . . 32
III.4.3 Predictions of the revised model . . . 33
III.4.4 Discussion . . . 36
III.5 Validation of the revised model . . . 39
III.5.1 Predictions . . . 39
III.5.2 Discussion . . . 41
III.6 General discussion . . . 41

IV A binaural model implementing an internal noise to predict the effect of hearing impairment on speech intelligibility in non-stationary noises 45
IV.1 Introduction . . . 45
IV.2 Model description . . . 46
IV.2.1 Original model developed for NH listeners . . . 46
IV.2.2 Extension of the model to HI listeners . . . 47
IV.2.3 Model evaluation . . . 48
IV.3 Results . . . 49
IV.3.1 Dataset involving NH and HI listeners . . . 49
IV.3.2 Dataset involving only NH listeners . . . 53
IV.4 Discussion . . . 54

IV.4.1 Improvements obtained with the new model . . . 54
IV.4.2 Further exploration of the internal noise implementation . . . 55
IV.4.3 Is the proposed internal noise concept in line with the literature ? . . . 57
IV.4.4 Hypotheses and limitations of the proposed model . . . 59
IV.5 Summary . . . 61

V The contribution of binaural unmasking to speech intelligibility in noise for normal-hearing and hearing-impaired listeners 63
V.1 Introduction . . . 63
V.2 General methods . . . 64
V.2.1 Listeners . . . 65
V.2.2 IPD sensitivity measurement . . . 65
V.2.3 SRT measurements . . . 66
V.2.4 Model . . . 68
V.3 Results . . . 69
V.3.1 IPD sensitivity measurements . . . 69
V.3.2 Effect of azimuth, reverberation, and masker type on SRTs . . . 70
V.3.3 Effect of presentation level on SRTs . . . 71
V.4 Discussion . . . 73
V.4.1 Influence of hearing loss . . . 73
V.4.2 Factors influencing the SRT . . . 75
V.4.3 Factors influencing binaural unmasking . . . 76
V.4.4 Predicting binaural unmasking . . . 76
V.5 Summary . . . 77

VI Concluding remarks and outlook 79
VI.1 Summary . . . 79
VI.2 Model strengths and weaknesses . . . 80
VI.3 Outlook . . . 81

References 83

A Description of the experiment (VL) conducted in Chapter III 93

B Comparison of the predictions from the models proposed by Lavandier et al. (2018) with the predictions from the model proposed in Chapter IV 97

C Comparison between the model developed in Chapter IV and the model with the revised jitters proposed in Chapter V 101

D Ethic approval letter 103

E Authorship contribution statement forms 105

Résumé en français 115
1 Introduction . . . 115
2 Etat de l'art . . . 116
2.1 Compréhension de la parole dans le bruit . . . 116
2.2 Modèle de compréhension de la parole dans le bruit . . . 117
2.3 Objectifs de la thèse . . . 117
3 Validation d'un modèle binaural de compréhension de la parole dans des bruits modulés en amplitude pour des auditeurs NH . . . 118
4 Un modèle binaural implémentant un bruit interne pour prédire l'effet de la perte auditive sur la compréhension de la parole dans des bruits modulés en amplitude 119
5 La contribution du démasquage binaural pour l'intelligibilité de la parole dans le bruit pour des NH et HI . . . 120
6 Conclusion . . . 121

List of Figures

II-1 Percentage of correct words as a function of SNR. 3

II-2 Illustration of the masking release when the masker is modulated in amplitude. The blue target signal is the same in both panels while the orange masker signal is either steady in amplitude (left panel) or modulated (right panel). In the right panel, masking is lower in the regions indicated by the arrows compared to the left panel, thus, improving target intelligibility. 4

II-3 Schematic illustrating better-ear listening: the blue target signal reaches the listener's ears without ILD while the orange masker presents an ILD. This difference in ILD between the target and masker signals produces a higher SNR at the left ear, so-called the better ear, allowing the listener to improve intelligibility. 5

II-4 Difference in root-mean-square power at the listener's ears produced by two noises presenting 8-Hz sinusoidal modulation and placed on both sides of the listener at $\pm 60°$. 6

II-5 Schematic illustrating binaural unmasking: the blue target signal arrives in phase at the listener's ears while the orange masker presents an ILD and ITD (left figures). According to Durlach's theory, the listener is able to internally equalize the masker signal by compensating for masker ILD and ITD at one ear (middle figures), and then partly cancel it to improve the internal SNR (right figure). Only the part that is similar at both ears is cancelled explaining why it remains target information and a masker residual at the end of the mechanism. 7

II-6 Replot of the data collected by Peissig and Kollmeier (1997). The target speech was in front of the listener (0°) while the steady-state SSN was moved around the listeners. 8

II-7 The standard audiograms proposed by Bisgaard et al. (2010) plotted in two panels according to their slope. Each marker style represents one hearing loss severity. . 9

II-8 Visual representation of the models presented in the current review as a function of the approach used to predict intelligibility as well as the listening scenarios in which they were tested (listening conditions and listener hearing profile). The hybrid models are placed over the two approaches considered to predict intelligibility. The non-intrusive models that require only the target+noise mixture as input are distinguished with blue boxes. The arrows between models highlight the fact they belong to the same family, and the double arrows are used to additionally show the binaural versions of the monaural models. The models used and further developed during the PhD are shown using a red font. The st(BE+BU)$_{\text{HI+HA}}$ (in faded red) is the long-term goal of the project, which is developing a model that is able to predict speech intelligibility in noise for HI listeners wearing hearing-aid devices. 24

II-9 Block diagram of the models that have been considered in the PhD project. The black font represents the model designed only for NH listeners Collin and Lavandier (2013) and the modifications associated with the extension to HI listeners are highlighted in grey Lavandier et al. (2018). 25

III-1 Block diagram of the original model (Collin and Lavandier, 2013), with the parameters tested in the present study highlighted in grey. 30

III-2 Mean SRTs with standard errors across listeners measured in CM1, involving 1 or 2 noises, steady-state or modulated by a 10-Hz square wave (50% duty cycle, modulated out-of-phase if they were two maskers), simulated as originating from different azimuths (0° and +105° or ±105° if there were two maskers) in an anechoic environment. The target was always at 0°. Model predictions are displayed as a solid line for the revised model and as a dashed line for the original model. Model performance statistics are displayed only for the revised model. . . 34

III-3 Mean SRTs with standard errors across listeners measured in CM2. The target was always presented at 0° in the presence of two noises placed on both sides of the listener (±105°). The noises were modulated out-of-phase by a square wave at 5 modulation rates (1, 2, 5, 10, 20 Hz). Three types of HRIR were involved (ILD+ITD, ILD-only, ITD-only). One reference condition involved a steady-state noise co-located with the target (modulation rate of 0 Hz). Model predictions for the separated configuration are displayed as a solid line for the revised model and as a dashed line for the original model. The predictions related to the co-located configuration are plotted using a cross and a plus sign for the revised and original model, respectively. Model performance statistics are displayed only for the revised model. 35

III-4 Mean SRTs with standard errors across listeners measured in CL1. The target was at 0.65 m in front of the listener in a lecture hall. The noise was placed at three distances (0.65, 1.25, 5 m), also in front of the listener. Four types of modulation were used for the noise (steady-state, 1-, 2- or 4-voice modulated). Model predictions are displayed as a solid line for the revised model and as a dashed line for the original model. Model performance statistics are displayed only for the revised model. 36

III-5 Mean SRTs with standard errors across listeners measured in CL4. The target was at 0.65 m in front of the listener in a meeting room. The single masker was always at 0.65 m but tested at two azimuths (0° and 25°). Three types of noise were involved (1-voice modulated, 2-voice modulated or steady-state). Two noises (steady-state or two 1-voice modulated) were tested in two configurations (0° or ±25°, 0.65 m). The revised and original model predictions are plotted as a solid and a dashed line, respectively. Model performance statistics are displayed only for the revised model. 37

III-6 Mean SRTs with standard errors across listeners measured in the present study (VL). The target was placed at 0.65 m, +25° from the listener (=target/near). The noise was steady-state (top panel) or 1-voice modulated (bottom panel). It was tested at two distances (near at 0.65 m and far at 5 m) and two azimuths (+25°/=target, -25°/ =target) in a room. Two types of BRIR were involved (natural BRIRs with ITD+ILD, SEIRs with no ITD/no tail). Solid lines present the revised model predictions, while dashed lines present the original model predictions. Model performance statistics are displayed only for the revised model. . 38

III-7 Mean SRTs with standard deviations across listeners measured by Ewert et al. (2017). SRTs are plotted as a function of the noise modulation type (steady-state, sinusoidal, 1-voice, 1-voice frequency incoherent). Each panel corresponds to a given type of HRIRs (ITD+ILD, ILD-only, ITD-only, IMBM, Independent Maskers). For the first four panels, two spatial configurations were tested : while the target was at 0°, the two noises were placed in front (at 0°) or on each side (±60°) of the listener (plotted in black circles and grey squares, respectively). For the last panel, two different HRIRs (natural and IMBM) at 0° were used to create the independent maskers represented with black circles and grey squares, respectively. The model predictions are plotted in solid black and grey lines in all panels. The model performance statistics across all conditions are indicated in the title, and the performances for each HRIR type are displayed in the corresponding panel. 40

IV-1 Block diagram of the proposed model. The original model designed only for NH listeners is presented in black (Vicente and Lavandier, 2020), the modifications associated with the extension to HI listeners are highlighted in grey. 46

IV-2 Mean SRTs with ±1 standard errors across NH listeners (blacks circles) and HI listeners (grey circles) measured by Rana and Buchholz (2016) as a function of masker type (SSN or VS). The two maskers were either co-located with the target in front of the listener ("co-loc", open symbols) or simulated on each side of the listener at ±90° ("separ", filled symbols). Mean predicted SRTs are displayed as downward triangles with the same filling and color patterns as the data. 51

IV-3 Mean SRTs (circles) with ±1 standard errors across NH listeners (left panel) and HI listeners (right panel) measured by Rana and Buchholz (2018a) at four overall masker sensation levels. Predicted SRTs are plotted with lines. The two maskers were VSs either co-located with the target in front of the listener ("co-loc", open symbols for data and dashed lines for predictions) or simulated on each side of the listener at ±90° ("separ", filled symbols for data and solid lines and predictions). 52
IV-4 Mean SRTs with ±1 standard errors across NH listeners (in blacks circles) and HI listeners (in grey circles) measured by Rana and Buchholz (2018b) as a function of spatialization method. The two maskers were VSs. Predicted SRTs are displayed as downward triangles with the same color pattern as the data. 53
IV-5 Individualized measured SRTs for HI listeners as a function of 4FAHL (grey squares) or individualized predicted SRTs (black circles). Only the experiment from Rana and Buchholz (2018b) is considered, each panel represents a spatialization method. The dashed and solid lines show the linear regressions between individualized measured SRTs and 4FAHL or individualized predicted SRTs, respectively. 54
IV-6 Individualized measured SRTs for HI listeners as a function of 4FAHL (right panel) or individualized predicted SRTs (left panel) across spatialization methods involved in Rana and Buchholz (2018b). The solid line within each panel shows the linear regression. 55
IV-7 Mean SRTs with ± 1 standard errors across listeners as a function of SSN position, measured by Lavandier et al. (2012). The conditions spatialized with BRIRs are plotted as black circles and those with SEIRs as grey squares. Predicted SRTs from the proposed model are displayed as solid lines and those predicted with the Lav12 model are plotted as dashed lines, with the same color pattern as the data. Model performance statistics are displayed only for the proposed model. 56
IV-8 Mean SRTs with ± 1 standard errors across listeners as a function of masker modulation depth, measured by Collin and Lavandier (2013) in their "variable" masker condition (circles). Target and noise signals were presented diotically. The SRTs predicted with the proposed model are displayed as downward triangles and those predicted with the Col13 model as upward triangles. Model performance statistics are displayed only for the proposed model. 57
IV-9 Internal noise level as a function of the model center frequency based on either an audiogram with 0 dB HL (NH listener, left panel) or the audiogram averaged across all HI listeners considered in this study (HI listener, right panel). In each panel, the three internal noise spectra are plotted as solid lines and shade of grey for an external noise level of 40, 60 and 80 dB SPL. 58
IV-10 Internal noise levels at 3 given center frequencies of the model as a function of the external noise level. In the left panel, the internal noise level is generated using an audiogram with 0 dB HL at any frequency and ear (NH listener). In the right panel, the audiogram averaged across all HI listeners considered in this study (HI listener) is used to compute the internal noise level. In the same panel, the influence of the NAL-RP amplification on the internal noise level is plotted as dashed lines. 59

V-1 Average audiograms across ears of the NH (left panel) and HI (right panel) listeners. Each listener is represented by a symbol and a single marker style. The black solid line represents the average audiogram across listeners in each panel. . 65
V-2 Mean IPD_{freq} with ±1 standard deviations across each listener group as a function of listener group. The little black circles show the individual data. 70
V-3 Mean measured (circles) and predicted (solid lines) SRTs with ± 1 standard deviation across listeners as a function of target azimuth. Each panel contains data and predictions for a listener group and a masker type. The SRTs collected with the different α values (1, 0.7, 0.2) are plotted as dark, medium and bright greys. The noise was simulated at 6.16 m, +16.4° from the listener in a virtual room with an absorption coefficient α and played at 60 dB SPL. The target was anechoic and at 2 m from the listener. 71

V-4 Same as Fig. V-3 but for the binaural unmasking advantages obtained by subtracting the SRTs at -90° and +40° from the SRTs at +20°. 73
V-5 Mean measured (circles) and predicted (lines) SRTs with ± 1 standard deviation across listeners as a function of masker level. The measured and predicted SRTs obtained with the SSN and VS maskers are plotted in grey and black, respectively. Each panel is for one listener group. The target was placed at 2 m, -90° from the listener while the (anechoic) masker was at 6.16 m, +16.4°. The SRTs predicted with the model with revised jitters are displayed as solid lines, while the dashed lines show the predictions of the Vic20 model. Performance statistics are displayed only for the model with revised jitters. 74

B-1 Mean SRTs with ±1 standard errors across NH listeners (black circles) and HI listeners (grey circles) measured by Rana and Buchholz (2016) as a function of masker type (VS or SSN). The two maskers were either co-located with the target in front of the listener ("co-loc", open symbols) or simulated on each side of the listener at +/- 90° ("separ", filled symbols). Mean predicted SRTs are displayed as downward triangles (proposed model) and squares (Lav18 models) with the same filling and color patterns as the data. 98
B-2 Mean SRTs (circles) with ±1 standard errors across NH listeners (left panel) and HI listeners (right panel) measured by Rana and Buchholz (2018a) at four overall masker sensation levels. Predicted SRTs are plotted with lines for the proposed model and with squares for the Lav18 models. The two maskers were VSs either co-located with the target in front of the listener ("co-loc", open circles, open squares and dashed lines for data, Lav18 models and proposed model, respectively) or simulated on each side of the listener at +/- 90° ("separ", filled circles, filled squares and solid lines for data, Lav18 models and proposed model, respectively). 99
B-3 Mean SRTs with ± 1 standard errors across NH listeners (in black circles) and HI listeners (in grey circles) measured by Rana and Buchholz (2018b) as a function of spatialization method. The two maskers were VS. Predicted SRTs obtained with the proposed model and Lav18 models are displayed as downward triangles and squares, respectively, with the same color pattern as the data. 100

C-1 Same as Fig. V-5 but for audibility advantage obtained by subtracting the SRTs at 50 and 60 dB SPL from the SRTs at 40 dB SPL. 101

List of Tables

II.a Summary of the models described in the manuscript, each line is allocated to one model. "Listening" indicates the listening condition in which the model can be applied. "Input" indicates the type of input the model requires (S = speech alone, N = noise alone, S+N = noisy target/target+noise mixture/distorted target, $S_{ILD+ITD}$ = target ILD and ITD, IR_S/IR_N = binaural room impulse response of the speech/noise location, respectively). "Approach" indicates the metric used by the model to predict intelligibility (SNR_{AFD}/SNR_{MD} = SNR in the auditory frequency/modulation domain, respectively, Corr = correlation between both inputs, Glimpse = target glimpse, ASR = automatic speech recognizer). "HI" indicates whether (✓) or not (×) the model can account for the effect of hearing impairment. "HA" indicates whether (✓) or not (×) the model has been tested on conditions involving non-linear hearing-aid processing. The sign ∼ (in the last two columns) means that the model is likely able to predict the effect of the HI or HA but it is yet to be confirmed. The presentation of the models is in the same order as in the text. The horizontal lines within the table shows the different subsections in the text. 14

III.a Summary of the experimental designs used to test the model parameters. The superscript 'T' indicates the target's distance and azimuth and defines the co-located condition. The last column indicates the nature of the binaural cues available in the tested signals. 29

III.b First order indices, their sum and the second order index between Ceiling and BE for all experiments. The highest first order index for a given experiment is displayed in bold. Only the main interaction (between Ceiling and BE) is displayed here. 33

III.c Performance statistics of the original (Orig.) and revised (Rev.) model. $MeanErr$, $RMSErr$ and $MaxErr$ are computed in dB. The experiment of Ewert et al. (2017) was only used to validate the revised model. 34

V.a Summary of the experimental design: each column represents a tested factor. Subset I contains the conditions investigating the effect of the room and the target azimuth on binaural unmasking. Subset II contains the conditions assessing the effect of level on binaural unmasking. The two SRTs measured when the target was placed at -90° with the anechoic noises (SSN or VS) played at 60 dB SPL were common to the two subsets. 67

V.b Summary of the significant interactions and their related Tukey pairwise comparisons for the subset of SRTs measured at 60 dB SPL (see Subset I in Table V.a). The first column set a condition, the second a contrast and the third gives the significance level. For instance, the first comparison indicates that for both HI and NH listeners, the SRTs are significantly higher when the target was masked by a SSN compared to a VS ($p < 0.0001$, for each group). 72

V.c Same as Table V.b but for binaural unmasking advantages measured at 60 dB SPL. 73

V.d Same as Table V.b but for SRTs measured at 40, 50 and 60 dB SPL (see Subset II in Table V.a). 75

B.a Comparison of performances between the Lav18 models and the model proposed in Chapter IV. 99

C.a Performance statistics of the Vic20 model and the model with revised jitters (Rev.). The row "All" reports the performance statistics computed with the predicted SRTs across all conditions. BU_{Adv} and Aud_{Adv} stand for binaural unmasking advantage and audibility advantage, respectively. 102

I Introduction

Hearing loss affected 850 million people in 1990, while in 2019, 1.57 billion people (about 20% of the world population) had at least a mild hearing loss (GBD 2019 Diseases and Injuries Collaborators, 2020). This rising number is due to the population growth as well as the improved life expectancy (Davis and Hoffman, 2019) because hearing loss increases with age. This results in a higher number of hearing-impaired people per 100,000 (from 15,890 in 1990 to 20,310 in 2019). The cost of unaddressed hearing loss is estimated in the range €620-660 billions (World Health Organization, 2017). This includes health-care expenditure related to hearing loss (except the costs of hearing-aid devices), productivity loss (e.g., due to unemployment) and societal/intangible costs. A recent literature review (Nordvik et al., 2018) reported that acquiring hearing loss in adult years lowers quality of life and is a risk factor for distress due to, for instance, social isolation. Oppositely, people who are suffering from hearing loss from younger ages better adapt to their impairment but present a lower education level. Hearing-aid devices can partly alleviate the effect of hearing loss, by restoring intelligibility and thus quality of life. In these circumstances, hearing loss is then a societal and economical burden that will increase in the next decades.

In order to reduce the burden of hearing loss by providing efficient hearing-aid devices, it is necessary as a first step to understand what aspects of hearing impairment affect the ability of hearing-impaired people to interact with our society. Especially, social isolation observed in hearing-impaired listeners suggests that they struggle with conversing in noisy environments such as restaurant, food court, pub... This is referred to as "cocktail-party" problem or situation by Cherry (1953). Experimental research has been conducted to better understand the auditory mechanisms that contribute to speech intelligibility in noisy environments for normal-hearing and hearing-impaired listeners. This has allowed to test, verify and design theories that have been further challenged by developing speech intelligibility models. Experimental and modelling approaches complement one another. Experimental data are required to design a model rationale while this model either verifies the rationale if it describes the effects observed in the data, or helps to revise this rationale if the predictions do not fit the entire dataset.

The outcomes of these researches suggest that a listener is able to segregate the target speech from the competing sources, so-called maskers, when they are spatially separated by using auditory mechanisms based on interaural level and time differences of the signals. However, these mechanisms are degraded when a listener suffers from hearing loss so that speech intelligibility can be greatly reduced. A number of speech intelligibility models from the literature are able to predict the spatial benefit observed for normal-hearing listeners when the speech is spatially separated from the maskers (as in cocktail-party situations). Few of them consider some aspects of hearing loss, others can account for the effect of hearing-aid processes but none of these models provide accurate predictions for an individual hearing-impaired listener wearing hearing aids in realistic environments.

That is why the motivation for this project is to develop a binaural speech intelligibility model that can account for the effects related to hearing loss in complex, realistic scenarios, including reverberation and competing masking sources. This model would help to better understand what aspects of hearing loss as well as hearing-aid features impact speech intelligibility. The long-term outcome is to design better hearing-aid devices (or signal processing features) that allow hearing-impaired listeners to communicate in challenging noisy conditions as normal-hearing listeners, which would improve their social interactions and then would reduce the burden of hearing loss.

Modelling cocktail-party situation can be complex because it involves various moving sources, interlocutors, hearing loss profiles, masking sources... All those factors cannot be addressed in

one PhD, hence, the framework is here limited to one spatially fixed target speech masked by fixed noise sources and some auditory mechanisms for both normal-hearing and symmetrically, mild to moderately severe hearing-impaired listeners. Noise maskers produce an energetic masking that results from the overlap between target and masker in the time, frequency and modulation domains. Speech maskers are not considered despite their presence in cocktail-party situations because they introduce an additional masking compared to noise maskers. This additional masking, so-called informational masking, is related to the similarity between target speech and speech masker as well as the target uncertainty (e.g., which source to listen to). The rationale is to isolate the perceptual mechanisms associated with speech intelligibility in cocktail-party situations using energetic noise maskers as a first step, and then extend the model to (informational) speech maskers in future projects.

The structure of the PhD emphasizes that each chapter is a study on its own and it addresses particular auditory mechanisms. The manuscript starts with Chapter II that aims to provide the required knowledge to understand the auditory mechanisms involved in speech intelligibility in noise for normal-hearing and hearing-impaired listeners as well as a literature review of the models that predict intelligibility for these listeners. Chapter III is about the optimization of a binaural speech intelligibility model for normal-hearing listeners applied on datasets testing auditory mechanisms in isolation and combination, thus, hearing impairment is discarded in a first place. Chapter IV investigates the influence of hearing loss on the auditory mechanism using interaural level differences of the signals to enhance speech understanding. Chapter V highlights the contribution of the auditory mechanism using interaural time differences to improve speech intelligibility for both normal-hearing and hearing-impaired listeners. Finally, Chapter VI summarizes the accomplished work highlighting how the results can be interpreted and used for further research.

II Background

II.1 Speech intelligibility in noise

Sections II.1.1, II.1.2 and II.1.3 focus on the auditory mechanisms related to speech intelligibility in noise for normal-hearing (NH) listeners. Section II.1.5 introduces the characteristics of hearing impairment and illustrates how the auditory mechanisms are degraded for hearing-impaired (HI) listeners.

II.1.1 Signal-to-noise ratio and intelligibility

The signal-to-noise power ratio (SNR) is an indicator of intelligibility. Figure II-1 illustrates a psychometric function, which represents the percentage of correct words as a function of SNR. Increasing the SNR allows a listener to better understand the target words. The SNR at which the listener understands half of the target words is called speech reception threshold[1] (SRT). One can observe two asymptotes in the lower and higher SNR ranges. This reflects the fact that there is an upper limit in SNR at which the entire words are fully understood, hence, increasing the SNR is not beneficial anymore. Oppositely, there is a lower limit in SNR at which all the words are unintelligible, thus, decreasing the SNR is not detrimental anymore. French and Steinberg (1947) demonstrated that intelligibility relies on the SNR per frequency band, i.e., the spectral differences between target and speech are relevant for intelligibility.

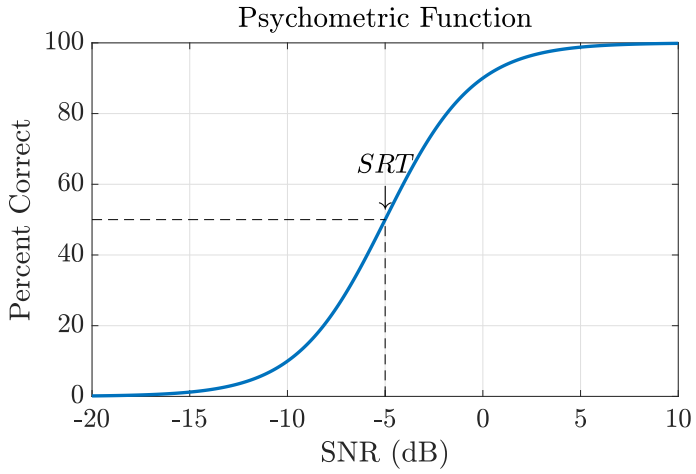


FIGURE II-1: Percentage of correct words as a function of SNR.

[1]The threshold can be set at any percentage but in this manuscript by default it is 50%.

II.1.2 Dip listening

Modulations in amplitude that occur in the envelopes of speech and masker signals lead to variations in the SNR across time. A listener can benefit from masker modulations by listening in the dips of the temporal envelope (Miller and Licklider, 1950; Fogerty et al., 2018; Festen and Plomp, 1990; Collin and Lavandier, 2013), i.e., when the short-term SNR is higher than the long-term SNR. A visual analogy is shown in Fig. II-2 to illustrate dip listening. A target sentence is plotted in blue and masked by a steady-state noise (left panel) or a modulated noise plotted in orange (right panel). The target signal can be barely seen when the noise is unmodulated while it appears better when the noise has modulation valleys (indicated by the black arrows).

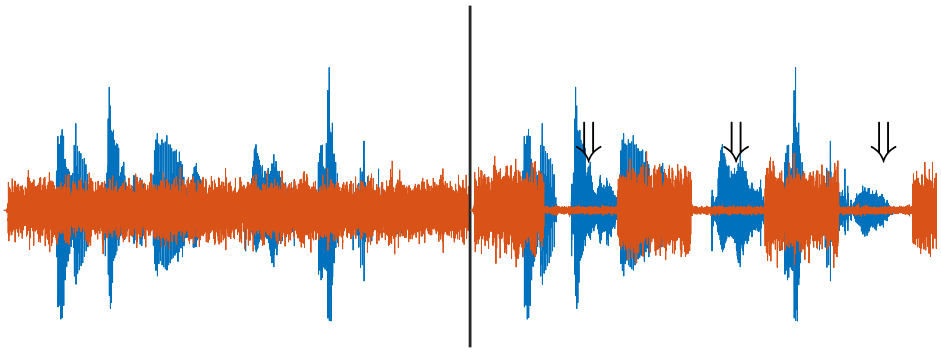


FIGURE II-2: Illustration of the masking release when the masker is modulated in amplitude. The blue target signal is the same in both panels while the orange masker signal is either steady in amplitude (left panel) or modulated (right panel). In the right panel, masking is lower in the regions indicated by the arrows compared to the left panel, thus, improving target intelligibility.

Miller and Licklider (1950) and Fogerty et al. (2018) investigated the effect of temporal fluctuation in the masker envelope using an unmodulated noise that they turned on and off following a square modulation, resulting in a "gated" noise. They showed that the highest speech intelligibility score occurred when the noise was modulated at a frequency of about 10 Hz. Fogerty et al. (2018) also tested the influence of random gating by shifting with a random phase each full period of the periodic-gating noises. The intelligibility scores obtained with the randomly gated noises were lower (on average across gating rates) suggesting that the listeners were able to predict the dip occurrence in the periodic noise to improve intelligibility. Collin and Lavandier (2013) also demonstrated the effect of the predictability of the masker dips using noises modulated by a broadband envelope of a speech signal. A "variable" condition was considered in which the modulated noise changed from one sentence to another while adaptively measuring the SRT. Oppositely, the "constant" condition involved the same modulated noise during one SRT measurement. The SRTs were lower under the constant condition, which suggests again that the listeners predicted the dip occurrence to enhance intelligibility.

II.1.3 Spatial release from masking

II.1.3.1 Binaural cues induced by the head

A signal of a source reaching the listener's ears presents an interaural level difference and an interaural time difference (ILD and ITD, respectively), due to head shadow. The head induces an ILD because it absorbs, reflects and diffracts the signal energy, thus, a source placed on one side of the listener produces a lower sound level at the ear on the opposite side. This attenuation varies across frequency, and because the human head tends to be acoustically transparent at low frequency due to its size between ears, the attenuation increases with frequency. Still considering a source on one side of a listener, an ITD is also created because of the difference in travel time

between the source and each listener's ear. This interaural delay can be also expressed in terms of an interaural phase difference (IPD), which is proportional to the product between ITD and frequency. For a sound with a temporal fine structure and an envelope (such as speech), the auditory system is able to detect an IPD up to 1,500 Hz using the temporal fine structure of the signal (e.g., Füllgrabe et al., 2017). An ITD in high frequency can be also detected if it is carried by the signal envelope (e.g., Akeroyd, 2006).

Thanks to ILD and ITD, intelligibility is improved when a speaker is spatially separated from the masking noise compared to a spatially co-located configuration, which can appear to be helpful in noisy environments (e.g., Bronkhorst and Plomp, 1988; Culling et al., 2004). This spatial release from masking (SRM) is commonly explained using two perceptual mechanisms: better-ear listening and binaural unmasking.

II.1.3.2 Better-ear listening

Better-ear listening uses differences in ILD between target and masker to improve intelligibility. Fig. II-3 is a schematic illustrating better-ear listening. The speaker (colored in blue) and the listener are facing each other while a noise source is located on the right side of the listener. The speaker signal reaches the listener's ears without ILD while the masker signal presents an ILD. This difference in ILD between both sources leads to a higher SNR at the left ear compared to the other. Better-ear listening is the listener's ability to use the ear with this better SNR to improve intelligibility.

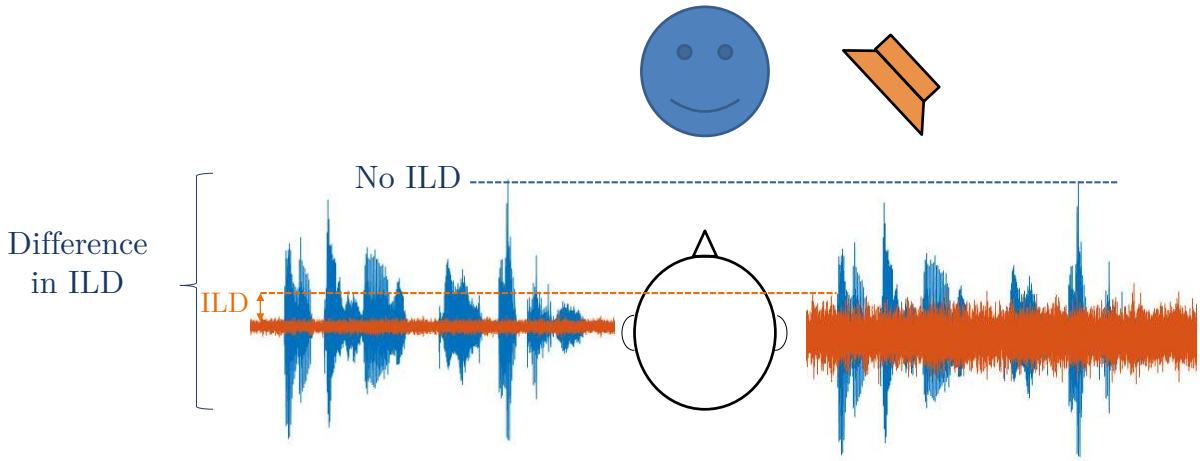


Figure II-3: Schematic illustrating better-ear listening: the blue target signal reaches the listener's ears without ILD while the orange masker presents an ILD. This difference in ILD between the target and masker signals produces a higher SNR at the left ear, so-called the better ear, allowing the listener to improve intelligibility.

In an environment with several masking noises, distributed around the listener, the better ear can change across time. This ability to switch back and forth to use one ear or the other to catch a target glimpse (Cooke, 2006) to improve intelligibility is called better-ear glimpsing. Figure II-4 shows the difference in power at the listener's ears produced by two noise maskers sinusoidally modulated (8 Hz) and located at $\pm 60°$. The root-mean-square power is computed per time frame using a 12-ms sliding square window. The masking energy is higher sometimes at the right ear sometimes at the left ear, thus inducing better-ear glimpsing considering a frontal target for instance (because the target energy would be the same at both ears).

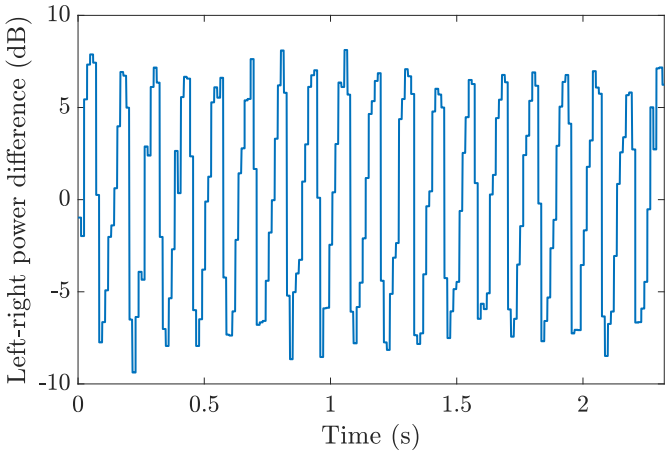


FIGURE II-4: Difference in root-mean-square power at the listener's ears produced by two noises presenting 8-Hz sinusoidal modulation and placed on both sides of the listener at ±60°.

It is quite unclear whether better-ear glimpsing relies on a binaural mechanism or two monaural mechanisms, one at each ear independent from each other. If the nature of better-ear glimpsing was binaural, the auditory system would compare the SNRs at the two ears to select the best one to improve intelligibility. Then, it would need to switch back-and-forth to follow the ear with the best SNR. For the monaural case, the auditory system would benefit of the SNRs at both ears simultaneously. The binaural system appears to be sluggish compared to the monaural system, i.e., the responses of the binaural system to interaural cue changes are more sluggish than the responses of the monaural system to monaural cue changes. Thereby, few studies attempted to clarify the true nature of better-ear glimpsing by using this difference in time responses. Brungart and Iyer (2012) generated an ideal monaural better-ear mask (IMBM) with binaurally modulated maskers. An IMBM simulates perfect better-ear glimpsing — but without binaural cues — as if a listener can switch instantly back-and-forth across time-frequency units from one ear to the other to catch a target glimpse. To create an IMBM, a binaural masker is decomposed in time-frequency units using a time frame that reflects the time resolution of the monaural system. Then, within each unit, the ear that receives the less energy is selected and the IMBM is generated gathering these energy-lowered time-frequency units. Brungart and Iyer (2012) showed that the intelligibility scores were not different between the IMBM condition and the natural binaural listening condition, thus suggesting that better-ear glimpsing relies on two monaural mechanisms. On the other hand, Culling and Mansell (2013) investigated the sluggishness related to binaural unmasking and better-ear glimpsing with modulated noises placed on both sides of a listener. The noises were modulated by a square wave at different rates. To force the listener to switch back-and-forth from one ear to the other, the noises on both sides of the listener were modulated out-of-phase. Culling and Mansell (2013) showed that better-ear glimpsing was subject to binaural sluggishness and concluded that better-ear glimpsing was a binaural mechanism. Ewert et al. (2017) measured SRM with modulated noises using (amongst other conditions) a spatialization method that involved only natural ILD and no ITD as well as a method to generate IMBM. The measured SRMs were higher for the ILD-only condition than for the IMBM conditions suggesting again that better-ear glimpsing advantage was limited without ILD information.

II.1.3.3 Binaural unmasking

Binaural unmasking is a mechanism that utilizes differences in ITD to improve speech intelligibility. It is explained and modelled in the equalization-cancellation (EC) theory from Durlach (1963, 1972) and a schematic is provided in Fig. II-5 to illustrate it. Durlach assumed that a listener is able to internally equalize the masking signals at the ears – by applying a gain and a delay at one ear to compensate for the masker ILD/ITD – and partially cancel this masker by subtracting the signals at one ear from the signals at the other ear to improve the internal SNR

and consequently speech intelligibility. The efficiency of the cancellation stage depends on the masker's interaural coherence at the listener's ears (see the noise residual after the cancellation stage in Fig. II-5), which reflects how similar the masking signals are at the ears.

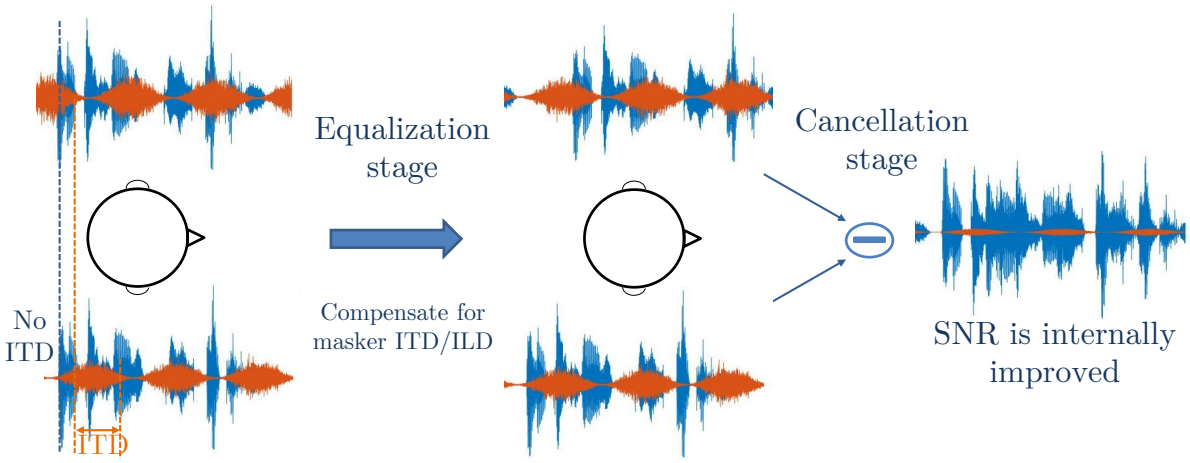


Figure II-5: Schematic illustrating binaural unmasking: the blue target signal arrives in phase at the listener's ears while the orange masker presents an ILD and ITD (left figures). According to Durlach's theory, the listener is able to internally equalize the masker signal by compensating for masker ILD and ITD at one ear (middle figures), and then partly cancel it to improve the internal SNR (right figure). Only the part that is similar at both ears is cancelled explaining why it remains target information and a masker residual at the end of the mechanism.

The advantage due to binaural unmasking for speech intelligibility in noise is up to 6 dB (Culling et al., 2004; Lavandier and Culling, 2010; Goverts and Houtgast, 2010; George et al., 2012). It increases with the spatial separation of the sources, the noise interaural coherence and the presentation level of the stimuli; but decreases when modulation in the masker envelope occurs. Bronkhorst and Plomp (1988) showed that the influence of the source azimuth difference is maximal between 0 and 30° and then the binaural unmasking benefit reaches a plateau, which was latter confirmed by Lavandier and Culling (2010). Licklider (1948) demonstrated that most of the intelligibility score variability occurs for a noise interaural coherence ranging from 0.75 to 1 (1 corresponding to the highest intelligibility). This was latter confirmed by Robinson and Jeffress (1963) using detection of tones in noise. They demonstrated that decreasing the noise interaural correlation from 1 to 0.95 can decrease the binaural advantage from 15 to 10 dB. This very steep decrease in binaural unmasking advantage when the noise starts to be slightly decorrelated across ears was also reported by Lavandier and Culling (2010) using a speech-in-noise paradigm.

II.1.3.4 Individual and combined contributions of the binaural cues

The maximum advantage due to SRM is about 10-12 dB (Culling and Lavandier, 2021; Peissig and Kollmeier, 1997; Culling and Mansell, 2013; Beutelmann and Brand, 2006; Bronkhorst and Plomp, 1988). Figure II-6 re-plots the data collected by Peissig and Kollmeier (1997) to illustrate the relationship between SRM and masker azimuth. The target speech was in front of the listener while the steady-state speech-shaped noise (SSN, noise with the same spectrum as the target speech) was placed at different azimuth around the listener. The highest SRMs were measured when the SSN was placed at +105° or -105°. A local minimum was observed for a masker placed at +90° or -90°, which was due to the appearance of constructive interference at

the opposite side of the head (Jelfs et al., 2011), leading to a higher masking sound level (Peissig and Kollmeier, 1997). The local minimum observed at 180° was due to a lower head-induced ILD and ITD (Bronkhorst and Plomp, 1988).

When there are more than one masking noise, their spatial configuration around the listener substantially affects intelligibility. The influence of the head shadow is reduced when there is at least one masker on each side of the target, thus limiting the advantage due to better-ear listening (Peissig and Kollmeier, 1997; Culling et al., 2004; Lavandier et al., 2012) and then SRM. The SRM observed with one steady-state noise is higher than with one modulated noise, likely because the modulated noise produces less masking over time (Hawley et al., 2004; Culling and Mansell, 2013) as it is illustrated in Fig. II-2 (there is no SRM when there is no masking).

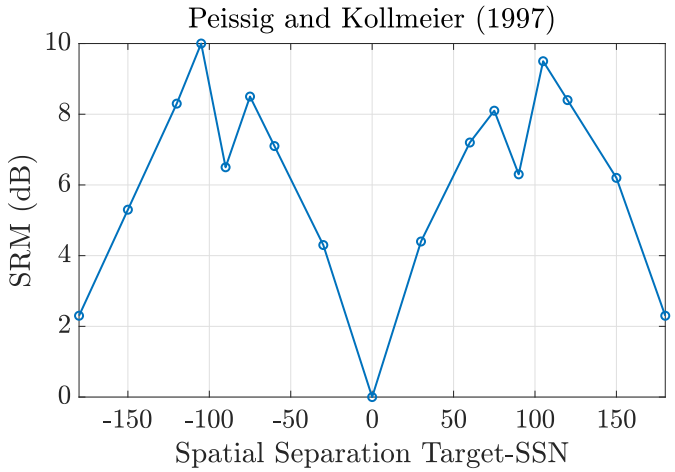


FIGURE II-6: Replot of the data collected by Peissig and Kollmeier (1997). The target speech was in front of the listener (0°) while the steady-state SSN was moved around the listeners.

Some studies investigated the contributions in isolation and/or in combination of better-ear listening and binaural unmasking to SRM. The rationale is to study whether the better-ear listening and binaural unmasking contributions interact and if one contribution is higher than the other. Bronkhorst and Plomp (1988) showed using one steady-state SSN that the combined contribution of better-ear listening and binaural unmasking was lower than the sum of two individual contributions suggesting that they interact. They also reported a higher contribution of better-ear listening than binaural unmasking on average across tested conditions. Culling and Mansell (2013) tested the influence of noises modulated by square waves at different rates and symmetrically placed on both sides of the target. They found that the individual contribution of better-ear listening was higher than the one of binaural unmasking. The sum of the two individual contributions was approximately equal to the combined contribution of the two mechanisms. Ewert et al. (2017) measured the SRM obtained with 6 masker types (included two speech maskers) symmetrically placed on both sides of the frontal target. They found that the binaural unmasking advantage was lower than the better-ear listening benefit on average across conditions and their contribution did not simply add up, i.e., the total benefit was lower than the sum of the benefits of the two individual contributions. These divergent findings might be led by the fact that the tested conditions are different across studies. However, the above conclusions agree on the fact that the individual contribution of binaural unmasking for intelligibility appears not to be higher than the individual contribution of better-ear listening.

II.1.4 Effect of room reverberation

Room reverberation can be quantified by the reverberation time that is the time for which the sound level in a room decays by 60 dB SPL after stopping any sound source in the room. Other metrics are also used to quantify room reverberation such as the ratio between the energies

of the early reflections and late reflections, which is typically called clarity or C_{50} (ISO 3382-1, 2009). This metric considers that the combined energy of the early reflections is useful for speech intelligibility because they are integrated with the direct sound and the energy of the late reflections is detrimental for speech intelligibility because they mask the direct sound. The fact that the target can be masked by itself due to reverberation is called target smearing (Rennies et al., 2011; Leclère et al., 2015).

Room reverberation distorts signal characteristics to such an extent that it can negatively affect auditory mechanisms such as dip listening, better-ear listening, and/or binaural unmasking. Reverberation fills the masker gaps due to temporal smearing, which decreases masker modulation depth and thus dip listening advantage (Collin and Lavandier, 2013). The multiple reflections due to reverberation tend to travel around the listener's head and reduce the head-related ILD and thus better-ear listening (Lavandier et al., 2012). Room reverberation also distorts the signal spectrum, which can create variations in SNR across frequency that can improve or deteriorate intelligibility due to differences in spectrum between target and masker (Lavandier and Culling, 2010). Binaural unmasking advantage can be degraded by reverberation because it reduces the noise interaural coherence at the listener's ears, which lowers the efficiency of the cancellation stage (Lavandier and Culling, 2008, 2010).

II.1.5 Effect of hearing impairment

II.1.5.1 Characterisation of hearing impairment

Hearing impairment is commonly known to induce an elevation of hearing thresholds resulting in reduced audibility, which leads to difficulty to hear soft sounds. A hearing threshold is estimated by playing a pure tone and varying its level to adaptively converge to the detection threshold of the signal. This value is expressed in dB hearing level (dB HL) representing the difference in dB SPL between a reference (designed with a large number of listeners with no hearing complaint) and the actual listener. This measure is done for different frequencies and for each ear because hearing loss can be frequency-specific and asymmetric. Figure II-7 represents standard audiograms proposed by Bisgaard et al. (2010) who analyzed a database of 28,244 audiograms. Hearing loss increases with age (age-related hearing loss or presbycusis, Spoor, 1967) especially at high frequency. Furthermore, noise exposure also induces hearing loss (Passchier-Vermeer, 1974) that is typically highest around 4 kHz, which can explain the profiles of the standard audiograms.

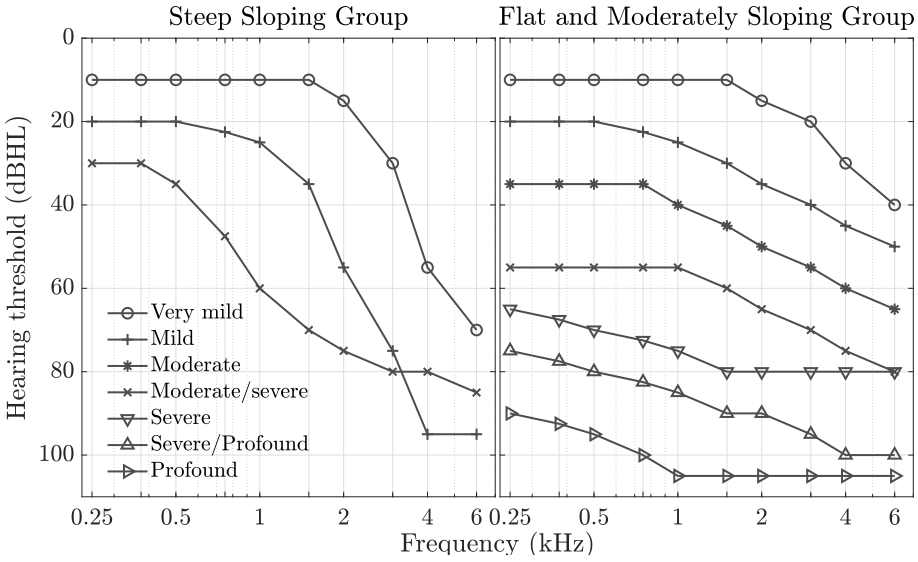


FIGURE II-7: The standard audiograms proposed by Bisgaard et al. (2010) plotted in two panels according to their slope. Each marker style represents one hearing loss severity.

Limited audibility has a direct effect on the target intelligibility, even in quiet, because it must be audible to be understandable. In addition, hearing threshold elevation also induces a loss of important speech and masker information, such as difference in ILD and ITD between the sources. This has a direct effect on the benefit provided by better-ear listening, since the speech signal at the ear with the better SNR may not be audible or only partially audible (Glyde et al., 2013). To say it in another way, head-induced ILD and the hearing thresholds for a (downward) sloping HI listener (see Fig. II-7) increase with frequency, thus limiting the benefit of better-ear listening. Reduced audibility also likely degrades perception of ITD because when a signal is not audible at one ear the auditory system cannot conclude on the signal ITD. Durlach et al. (1981) reported results from previous experiments that showed lower binaural unmasking advantage for asymmetric HI listeners, mostly because audibility was not equalized at the two ears.

Other deficits are also characteristic of hearing impairment, such as loudness recruitment in which changes in signal level typically result in larger perceived changes in loudness in HI than in NH listeners (e.g., Moore et al., 1999). Because of loudness recruitment, the range between hearing threshold and loudness discomfort (dynamic range) of HI listeners is smaller compared to NH listeners due to higher thresholds and similar or slightly higher — but not as much as the difference between hearing thresholds — loudness discomfort (Kamm et al., 1978). Moore et al. (1996) investigated the effect of recruitment on the perception of amplitude modulation in the masker envelope and showed that the perceived modulation depth was smaller for HI listeners, that likely degrades the mechanism of dip listening. In addition, Fitzgibbons and Wightman (1982) and Glasberg et al. (1987) measured detection of temporal gaps for NH and HI listeners and observed that HI listeners demonstrated a reduced temporal resolution meaning that they needed a longer gap in the signal to detect it. This deficit likely impacts further the ability to listen to the target in the masker dips.

Moreover, HI listeners show reduced sensitivity to the temporal fine structure of a signal (e.g., Moore, 2008). The latter has direct implications for ITD sensitivity in HI listeners (Füllgrabe and Moore, 2017). Whereas NH listeners are sensitive to changes in ITD up to a frequency of around 1.3 kHz, HI listeners show reduced sensitivity to ITD with a decrease in the upper cut-off frequency (King et al., 2014; Neher et al., 2011). Strelcyk and Dau (2009) and Neher et al. (2011) found a significant correlation between reduced ITD sensitivity and reduced intelligibility in conditions that maximized the difference in ITD between target and masker.

The auditory filters of HI listeners are broader than those of NH listeners (Glasberg and Moore, 1986), degrading the spectral analysis of the incoming signals. Speech and noise might be not differentiated because of a spectral smearing due to a wide overlap of the auditory filters

thus making speech segregation in noisy environment more challenging for HI listeners (Lesica, 2018).

II.1.5.2 Speech recognition in noise for HI listeners

Speech recognition, dip listening advantage and SRM are typically degraded by hearing impairment, which leads to more difficulties with understanding speech in noisy environments for HI listeners. This is also often confounded with the effect of aging on intelligibility (e.g., Glyde et al., 2011; Helfer et al., 2017) because hearing loss naturally increases with age and aging *per se* degrades auditory Schneider and Pichora-Fuller (e.g., 2001) and cognitive abilities (such as working memory, Kathleen Pichora-Fuller et al., 1995). The differences in intelligibility threshold between NH and HI listeners vary across studies, which depends on hearing loss, speech material, presentation level, type of masker, amplification, age difference between groups... A selection of some studies are provided here to illustrate this variability.

The benefit due to dip listening depends on the long-term SNR at which the sources are presented. This benefit decreases when the SNR increases (Jensen and Bernstein, 2019) because at high SNR a listener does not require dip listening to unmask the target because its level is high enough to be fully intelligible. Hence, for a fair comparison of performances between NH and HI listeners, only studies testing both groups at the same fixed SNRs are considered in this paragraph. Jin and Nelson (2006) measured dip listening advantage using modulated gated noise and amplifying the stimuli played to the HI listeners by half of their individual (frequency-dependent) hearing loss. The results highlighted 20 to 40% less speech recognition for the HI listeners. Phatak and Grant (2012) and Phatak and Grant (2014) demonstrated that HI listeners performed worse than NH listeners in using masker envelope modulation to recognize consonants while they showed equal performance as NH listeners for a vowel recognition task. The stimuli played to HI listeners were linearly amplified using the Cambridge formula to compensate partly for their hearing loss. Desloge et al. (2010) simulated hearing loss to NH listeners (i.e., simulated HI listeners) by adding a noise to the stimuli and then measured dip listening advantage using a gated noise. In this framework, HI listeners and simulated HI listeners obtained similar dip listening advantages, which means that audibility was the main factor degrading this advantage. Jensen and Bernstein (2019) designed a masker for each HI listener by removing any time-frequency unit that was below 10 dB above the hearing thresholds of the respective listener. Each HI listener was paired with a NH listener, who was tested with the same masker. Thereby, they demonstrated that HI listeners benefited of the modulation in a noise as much as NH listeners for low modulation rate. One general observation emerges from those studies, dip-listening advantage in HI listeners is always lower when gain is applied to the stimuli to attempt to restore audibility, however, hearing loss is actually not fully compensated. On the other hand, when audibility is degraded for NH listeners to mimic hearing impairment, dip listening advantage is equal across listener groups for low modulation rate. Hence, audibility is likely the main factor influencing dip listening followed by the difference in temporal resolution between NH and HI listeners.

Gelfand et al. (1988) differentiated the effect of hearing loss and aging by involving a group of young, middle-aged and old NH listeners (20-39, 40-54 and $\geqslant$55 years old, respectively) and a group of old HI listeners ($\geqslant$55 years old). The speech was predictable sentences placed at 0° or 90° and masked by a 12-talker babble noise located at either azimuth. The HI group presented significantly lower speech recognition compared to any NH group. The thresholds measured with the oldest NH group were significantly lower than those of the youngest NH group but there was a difference in hearing threshold that was not controlled between both groups. This makes the conclusion on the influence of age harder. Helfer et al. (2017) reviewed 4 experiments conducted in their lab and concluded that speech recognition declined for their oldest tested groups ($\geqslant$ 60 years old) but they did not try to separate the effects of age-related hearing loss and aging. Dubno (2015) presented the preliminary results of a large-scale study investigating the effect of aging on speech recognition over 25 years and hundreds of listeners. The author showed that even when controlling for natural threshold increase, speech recognition in 12-talker babble noise (SNR = 8 dB) declined with aging for non-predictable sentences. It declined also for predictable sentences but only for men. However, the author emphasized on the fact that the decline in speech recognition with aging is lower than the decline due to age-related hearing loss.

Bronkhorst and Plomp (1989) measured SRM for HI listeners using a frontal speech masked by one steady-state SSN placed at different azimuths. They applied a frequency-independent gain to the stimuli that was defined by the listener's hearing threshold and upper limit of loudness comfort. They showed that the SRMs measured with both symmetrically and asymmetrically HI listeners were lower compared to NH listeners mostly due to a loss of better-ear listening advantage. However, both HI groups showed also a decline in binaural unmasking advantage across conditions although the symmetrically HI listeners had the same binaural unmasking advantage as NH listeners at the largest tested azimuth separation (equal to 90°).

Glyde et al. (2011) reviewed 12 studies to investigate the effects of aging and hearing loss on SRM, eight of which considered noise maskers and comparison between listener groups (classified by age and/or hearing loss). Two studies did not control age-related hearing loss: Gelfand et al. (1988) reported no difference in SRM between the youngest and oldest NH groups while Kim et al. (2006) found a negative, significant correlation between age and spatial benefit. Four studies still found an effect of aging after controlling age-related hearing loss using a statistical method (Divenyi and Haupt, 1997; Divenyi et al., 2005) or a filtering method (Dubno et al., 2002; Murphy et al., 2006). On the other hand, Dubno et al. (2008) did not measure any difference between a younger and an older NH group using low-pass filtered stimuli to avoid the effect of age-related hearing loss. Finally, Arbogast et al. (2005) did not find any effect of hearing impairment for the conditions where speech was masked by noise so they did not further investigate the effect of aging. To sum up, Glyde and colleagues suggested that older HI listeners present reduced SRM but further studies need to clarify which factors affect SRM. It can be due to sensorineural hearing impairment, age-related cognitive changes or a combination of both. This spatial processing decline can be also due to aging and distinct from age-related cognitive issue or hearing impairment.

Best et al. (2013) compared SRM measured with sentences and speech-like modulated noise for age-matched, young NH and HI listeners. The stimuli were amplified using an individualized, frequency-specific linear gain following the National Acoustics Laboratories–Revised Profound (NAL-RP) prescription formula (Dillon, 2012, Chap. 10, pp 290-297) prescription formula to partly compensate for hearing loss. The young HI group demonstrated lower spatial benefit and this decline can be explained by the individual audiogram. This means that if the signals were amplified in a way to fully compensate for hearing loss, there would not be any difference across listeners. This outcome is supported by the result of Rana and Buchholz (2018a) and even extended to situations in which mean age of the groups differed (e.g., young NH versus old HI). They individually equalized audibility across frequency for all listeners, i.e., the sensation level was the same for all listeners. This resulted in different sound pressure level across ears and listeners, and substantial higher levels for HI listeners. They measured SRTs (and derived SRMs) using sentences and noise-vocoded speech masker played at different sensation levels and reported no significant difference between groups. This confirmed the above statement, there is no difference between listener groups when audibility is carefully equalized. However, Rana and Buchholz (2018a) had to be very careful with their stimulus design to avoid loudness complaints from HI listeners, which means that this kind of amplification cannot be considered for practical application.

There is a lack of consensus between the studies that might be related to differences in group age, masker type, filtering or amplification method (in other word, audibility) and/or speech material as suggested by Ahlstrom et al. (2014). However, some general trends can be observed. Aging degrades speech recognition but not systematically SRM nor dip listening advantage, while hearing loss degrades the three. Audibility seems to be the leading factor influencing SRM (Best et al., 2013), but when it is carefully equalized there is no difference between listener groups (Rana and Buchholz, 2018a), such as for the dip listening advantage (Jensen and Bernstein, 2019). Future studies need to provide more results about the effect of hearing loss and aging in isolation and combination in order to further test these suggested conclusions.

II.1.5.3 Hearing aids

Hearing aids are devices that aim to improve speech intelligibility for HI listeners. However, not all the characteristics of hearing impairment can be addressed and fully compensated for. They provide amplification of the sounds that arrive at the user's ears depending on his/her audiogram. However, the gain provided by the amplification cannot fully compensate for hearing

threshold elevation due to loudness discomfort. For instance, when Rana and Buchholz (2018a) individually equalized audibility across frequency for their listeners, the reference 0 dB sensation level (dB SL) was the individual SRT in quiet. It turned out that over the 10 HI listeners, only 9, 6 and 1 were able to be tested at 10, 20 and 30 dB SL, respectively, without complaining of loudness discomfort. This illustrates well the fact that providing gains equal to the listener's hearing thresholds cannot be considered in practice.

The first generations of hearing aids provided frequency-specific linear gains (such as the NAL-RP prescription gain, Dillon, 2012, Chap 10, pp 290-297) that compensated partly for hearing loss and amplified sounds regardless of their level. However, the users complained about loudness in noisy environments with such gains. Hence, wide dynamic range compression (WDRC) is preferred to make soft sounds audible while, at the same time, it reduces amplification for loud sounds to prevent loudness discomfort (e.g., Dillon, 2012, Chap 6, pp 180-189 or Kates and Arehart, 2005). Non-linear gain prescription formulas, such as NAL-NL2 (Keidser et al., 2011) or CAMEQ2-HF (Moore et al., 2010), are used to prescribe the amounts of amplification and compression that are applied in a hearing aid as a function of frequency. Thereby, the prescribed amplification is designed to maximize speech intelligibility (in quiet) while adjusting the perceived loudness to that of a NH listener. Hearing aids are also equipped with a range of more advanced signal processing features including beamforming (Dillon, 2012, Chap 7, pp 198-225), (spectral) noise suppression (Kjems et al., 2009), frequency lowering (Souza et al., 2013), and more (Dillon, 2012, Chap 8, pp 226-253 or Kollmeier and Kiessling, 2018). Beamforming can be seen as a spatial weighting that favours the near frontal sources by attenuating the sources in the other regions in space. Spectral noise suppression aims to lower the sound level of the signal for the frequency regions in which the SNR is low. Frequency lowering is a process that shifts the signal spectrum towards low frequency because the majority of hearing loss profiles presents lower loss in low frequency (see Fig. II-7).

II.2 Speech intelligibility models

This section provides a brief overview of the plurality of models that have been developed in the literature and emphasizes on the different approaches to model intelligibility (i.e., decision metrics). The models described here are those taking into account hearing impairment, binaural listening, and/or non-linear hearing-aid processes. The models are presented in three subsections: one focusing on monaural models, another one on binaural models for NH listeners only, and the last one on binaural models for HI listeners and/or non-linear hearing-aid processing. Within each subsection, the presentation of the models follows a chronological order. In the end, this section helps to show the contribution of the PhD project to the field of research.

For an objective comparison, some performance statistics are reported for each model, they are either found in the original publication or computed for the present review. These common model performance statistics are the Pearson's correlation coefficient (r) computed between data and predictions, the mean absolute error ($MeanErr$) computed as the mean across conditions of the absolute difference between data and predictions, the root-mean-square error ($RMSErr$) and the maximal absolute error ($MaxErr$).

Table II.a provides an overview of the models that are reviewed below. They are presented in the same order as in the text. The first column indicates the model acronym from the original publication or for few of them the acronym is defined in the current review. The second column indicates if the model is designed for monaural or binaural listening. The third column provides the required input(s) for the predictions. The input can be the clean speech (S), the noise alone (N), the noisy target/target+noise mixture/distorted target (S+N), the target ILD and ITD ($S_{ILD+ITD}$) or the binaural room impulse response of the speech/noise location (IR_S/IR_N). The fourth column gives the approach (or rationale) that the model uses to predict intelligibility. The models listed below based their intelligibility predictions on the SNR in the auditory frequency domain (SNR_{AFD}), the SNR in the modulation domain (SNR_{MD}), the correlation between both inputs (Corr), target glimpses (Glimpse) or using an automatic speech recognizer (ASR). The fifth column shows by a check mark if the model was tested on a dataset involving HI listeners. The last column shows also by the use of a check mark if the model was tested on experimental conditions involving non-linear hearing-aid processing. In the two last columns, there is also the sign $\sim$ that refers to the fact that the model is likely able to predict the considered effect (HL

or HA column) because a previous model version is able to, but the backward compatibility is yet to be confirmed.

II.2.1 Monaural models for hearing-impaired listeners or hearing-aid processing

II.2.1.1 The speech transmission index

Historically, the monaural modulation-based speech transmission index (STI) was developed by Steeneken and Houtgast (1980) because they highlighted that the articulation index (forerunner of the SII, see Sec. II.2.1.2) was able to predict distortion in the frequency domain (interfering noise or band-pass filtering) but was irrelevant in presence of non-linear distortions (e.g., peak clipping) or reverberation. For instance, the STI is more relevant to predict the temporal smearing of the target, i.e., the fact that the target is masked by itself due to late reverberation. The rationale of the STI is to evaluate per frequency band the reduction of the temporal modulations in the envelope of the speech signal due to distortion — which is estimated by the modulation transfer function — in order to derive an intelligibility metric.

Goldsworthy and Greenberg (2004) reported that neither the original implementation of the STI that uses an artificial speech-like signal as target nor the several speech-based STI from the literature that use speech signal as target were suitable to predict non-linear processing such as spectral subtraction or envelope clipping. Thereby, the authors proposed 4 new STI versions that could be able to predict such non-linear processes, however, they did not validate their approaches with experimental data.

II.2.1.2 The speech intelligibility index

The speech intelligibility index (SII, ANSI S3.5, 1997) is a SNR-based monaural index (between 0 and 1) that predicts speech intelligibility in noise. The SII is a improved version of the articulation index (Kryter, 1962). It is obtained by integrating the SNRs (limited in the range of [-15;15] dB) across frequency bands using a SII weighting that favours the frequency regions relevant for intelligibility as well as a frequency-specific distortion factor that degrades speech intelligibility when target levels are higher than a reference. The hearing loss of a listener is modelled using an internal noise, which is realized by adding the listener's pure-tone audiograms averaged across ears to a reference internal noise spectrum level (the internal noise level for a NH listener with 0 dB HL). For the computation of the SNR in each band, the higher level between internal and external noise is chosen.

The SII was thoroughly tested with HI listeners by Ching et al. (1998). Forty HI listeners were recruited with mild to profound flat or downward sloping hearing loss. The experiments consisted in sentence or consonant recognition in quiet. Band-pass and low-pass filtering were considered with different cutoff frequencies. They highlighted the importance of the distortion factor, otherwise speech intelligibility was overestimated at high sensation levels. The SII tended also to underpredict intelligibility scores at low sensation levels. The predictions could be further improved using an individual frequency-dependent proficiency factor, which reflects the listener ability to use audible speech in certain frequency regions. They showed that this factor for the region 2.8-5.6 kHz can be lower than or equal to 0 for severe HI listeners (>80 dB HL at 4 kHz). This means that those listeners did not benefit of the audible speech in this regions and could be even detrimental.

Rhebergen et al. (2010) showed that SII scores varied when using SNRs equal to SRTs even though by definition the number of correct words is fixed across SRTs, which must lead to equal SII scores across conditions. In addition, the inter-subject variability was higher for the SII scores than for the SRTs. To alleviate these discrepancies, Rhebergen et al. (2010) suggested to apply the cochlear compression to the stimuli before computing the SII scores.

The SII is not able to account for the effect of fluctuations in the masker envelope, thus, Rhebergen and Versfeld (2005) proposed a model (further tested by Rhebergen et al., 2006) that applies the SII per time frame and then averages the SII scores across time. However, this version was tested with HI listeners by Meyer and Brand (2013).

Model (article)	Listening	Input	Approach	HI	HA
STI (Steeneken and Houtgast, 1980)	Monaural	S & S+N	SNR_{MD}	×	∼
SII (ANSI S3.5, 1997)	Monaural	S & N	SNR_{AFD}	✓	×
CSII (Kates and Arehart, 2005)	Monaural	S & S+N	Corr	✓	✓
STOI (Taal et al., 2011)	Monaural	S & S+N	Corr	×	✓
mr-sEPSM (Jørgensen et al., 2013)	Monaural	N & S+N	SNR_{MD}	×	✓
SRMR (Falk et al., 2013)	Monaural	S+N	SNR_{MD}	✓	✓
HASPI (Kates and Arehart, 2014)	Monaural	S & S+N	Corr	✓	✓
mr-GPSM (Biberger and Ewert, 2017)	Monaural	N & S+N	SNR_{MD}+SNR_{AFD}	×	✓
$sEPSM^{corr}$ (Relaño-Iborra et al., 2019)	Monaural	S & S+N	Corr	×	✓
BSTI (Van Wijngaarden and Drullman, 2008)	Binaural	S & S+N	SNR_{MD}	×	×
STEC (Wan et al., 2014)	Binaural	S & N	SNR_{AFD}	×	×
B-sEPSM (Chabot-Leclerc et al., 2016)	Binaural	N & S+N	SNR_{MD}	×	∼
BiDWGP (Tang et al., 2016)	Binaural	S & N	Glimpse	×	×
G_{EC}+CSII (Mi and Colburn, 2016)	Binaural	S+N & $S_{ILD+ITD}$	Glimpse+Corr	×	×
PerG (Josupeit and Hohmann, 2017)	Binaural	S & N	Glimpse	×	×
stBSIM (Beutelmann et al., 2010)	Binaural	S & N	SNR_{AFD}	✓	×
combinedBSIM (Rennies et al., 2011)	Binaural	S & N & IR_S	SNR_{AFD}	∼	×
BSIM2020 (Hauth et al., 2020)	Binaural	S+N then S & N	SNR_{MD} then SNR_{AFD}	∼	∼
st(BE+BU) (Collin and Lavandier, 2013)	Binaural	S & N	SNR_{AFD}	×	×
B-SRMR (Cosentino et al., 2014)	Binaural	S+N	SNR_{MD}+ SNR_{AFD}	∼	∼
combined(BE+BU) (Leclère et al., 2015)	Binaural	IR_S & IR_N	SNR_{AFD}	×	×
$st(BE+BU)_{Aud}$ (Lavandier et al., 2018)	Binaural	S & N	SNR_{AFD}	✓	×
MBSTOI (Andersen et al., 2018)	Binaural	S & S+N	Corr	×	✓
FADE (Schädler et al., 2018)	Binaural	S+N	ASR	✓	✓

Table II.a: Summary of the models described in the manuscript, each line is allocated to one model. "Listening" indicates the listening condition in which the model can be applied. "Input" indicates the type of input the model requires (S = speech alone, N = noise alone, S+N = noisy target/target+noise mixture/distorted target, $S_{ILD+ITD}$ = target ILD and ITD, IR_S/IR_N = binaural room impulse response of the speech/noise location, respectively). "Approach" indicates the metric used by the model to predict intelligibility (SNR_{AFD}/SNR_{MD} = SNR in the auditory frequency/modulation domain, respectively, Corr = correlation between both inputs, Glimpse = target glimpse, ASR = automatic speech recognizer). "HI" indicates whether (✓) or not (×) the model can account for the effect of hearing impairment. "HA" indicates whether (✓) or not (×) the model has been tested on conditions involving non-linear hearing-aid processing. The sign ∼ (in the last two columns) means that the model is likely able to predict the effect of the HI or HA but it is yet to be confirmed. The presentation of the models is in the same order as in the text. The horizontal lines within the table shows the different subsections in the text.

II.2.1.3 The coherence-based speech intelligibility index

Kates and Arehart (2005) developed the coherence-based speech intelligibility index (CSII), which is an extension of the SII to account for the effects of clipping distortion and additive noise on intelligibility. The index compares the clean target speech signal to the distorted target speech signal (e.g., target with additive noise or clipped target) to estimate intelligibility. The procedure of the CSII is similar to the SII except that the SNR is replaced by the signal-to-distortion ratio. The signal-to-distortion ratio is estimated in each band (using ro-ex band-pass filters, Moore and Glasberg, 1983) by computing first the coherence between the clean target speech and the distorted target speech, then calculating the ratio between the coherence and 1 minus the coherence. Thereby, when there is no distortion (coherence equal to 1), the signal-to-distortion ratio goes to infinity while when distortion increases the signal-to-distortion ratio tends towards 0. Finally, the authors showed that the predictions could be improved by weighting differently three level regions, the mid level region (0 to 10 dB below the long-term RMS level of the speech signal) being the most important then the low level region (10 to 30 dB below the long-term RMS level of the speech signal) while the high level region (above the long-term RMS level of the speech signal) is discarded with a weight equal to 0. The listener's hearing loss is taken into account by the CSII in a same way as the SII, i.e., when hearing loss increases the signal-to-distortion ratio decreases so that intelligibility score decreases.

The CSII was validated in the original publication with an experiment involving NH an HI listeners. The effects of additive noise (at different SNRs) and clipping distortions were tested. The model accurately predicted the data averaged within each group ($r = 0.94$ for NH listeners and $r = 0.98$ for HI listeners) as well as for individual data collected with HI listeners (r ranging from 0.88 to 0.99). The prediction errors increased with the listener's hearing loss.

II.2.1.4 The short-time objective intelligibility measure

The short-time objective intelligibility measure (STOI) was proposed by Taal et al. (2011). This index takes as input the target speech and the distorted target speech. Both signals are decomposed in the time-frequency domain then the distorted envelope is normalized across time (to compensate for level differences between the signals) and clipped if the distortion is higher than a given level (in order to avoid an overwhelming effect of a time-frequency unit). Both envelope signals are compared by means of a correlation coefficient. The lower the distortion the higher the correlation, the higher the intelligibility. The coefficients are averaged across time and frequency to give a broadband value that can be compared to a speech intelligibility threshold using a mapping method. The STOI is extended (ESTOI) by Jensen and Taal (2016) in order to improve predictions for modulated noises by computing the spectral correlation rather than the temporal correlation between the signal envelopes.

The models were validated on a wide range of listening conditions including noise-reduction algorithms, different SNRs and noise types. The correlation r for the modulated noise condition was improved from 0.48 to 0.85 with the ESTOI (compared to the STOI). The correlation r considering the conditions with the noise-reduction algorithms was equal to 0.96 with the ESTOI.

II.2.1.5 The speech-based envelope power spectrum model

Jørgensen and Dau (2011) developed an alternative modulation-based approach to predict intelligibility for conditions involving non-linear processing. The input signals are the noisy speech and the noise, which are first decomposed by auditory frequency using a gammatone filterbank. The envelope of the resulting signals are extracted and then passed through a modulation filterbank to get the modulation envelope spectra. The ratios between signal+noise and noise alone ((S+N)NR) are estimated and combined first across modulation frequency then auditory frequency. Finally, the resulting value is converted into a speech intelligibility threshold using a mapping function.

Jørgensen et al. (2013) developed the multi-resolution speech-based enveloped power spectrum model (mr-sEPSM) that is an extension of the model proposed by Jørgensen and Dau (2011) and allows to take into account the fluctuation in the noise envelope. This model version estimates the modulation spectra of the signals using a time-frame decomposition, with a frame

duration that is equal to the inverse of the modulation center frequency. Thereby, the (S+N)NRs are averaged across time, combined across modulation and auditory frequency to be finally converted into an intelligibility threshold. The model accurately predicted conditions involving modulated noises ($RMSErr$ = 0.8 dB), steady-state SSN with reverberation ($RMSErr$ = 0.6 dB) or spectral subtraction ($RMSErr$ = 1.3 dB).

II.2.1.6 The speech-to-reverberation modulation energy ratio

Falk et al. (2013) developed the speech-to-reverberant modulation energy ratio (SRMR) is a monaural non-intrusive (i.e, considering only the target+noise mixture) metric that analyzes per frequency band the modulation in the envelope of the target+noise mixture. The mixture is passed through a gammatone filterbank (filters with approximately the same shape as auditory filters) and the envelopes of the output signals are extracted. A modulation filterbank is used to get the modulation envelope energy. Finally, a low-to-high modulation energy ratio is computed considering modulation frequencies below 20 Hz as speech modulation, while considering the frequencies above 20 Hz as modulation due to noise or room acoustics. A listener's hearing loss in the model is considered within the filterbank by broadening the filters (Glasberg and Moore, 1986).

The model was validated using a dataset collected with hearing-impaired listeners. The influences of reverberation, SNR, hearing-aid setting and noise type were investigated. Over the 40 conditions, the correlation r was equal to 0.84 and the $RMSErr$ was equal to 9.2%.

II.2.1.7 The hearing-aid speech perception index

Kates and Arehart (2014) developed the monaural hearing-aid speech perception index (HASPI) that predicts the effect of hearing-aid processing on intelligibility. The model inputs are the reference target signal (not processed) and the target signal modified by any process (additive noise, amplification, hearing-aid processing,...). Then, the reference signal is passed through an auditory periphery model for an ideal listener with no hearing loss. The same model is used for the processed signal, but taking into consideration the listener's hearing loss. The envelopes and temporal fine structures of the reference and processed signals are extracted and compared to give an index that reflects how similar they are (in a same vein as the CSII).

The model was tested on datasets involving NH and HI listeners, additive noise, clipping distortion, noise suppression, frequency lowering and noise vocoding. A linear gain was applied to the stimuli for the HI listeners to compensate partly for their hearing loss. The metric predicted well the trends observed in the data (r = 0.97 combining both groups). The authors also reported that the inter-subject variability was not as well captured by the model (r = 0.91 for the HI listeners and r = 0.93 for the NH listeners). A statistical analysis showed that the correlation r was either significantly better compared to the CSII or not different.

II.2.1.8 Combined models with the mr-ESPM

Relaño-Iborra et al. (2016) proposed a model that combines the auditory processing of the mr-sEPSM with a correlation metric (sEPSM$^{\text{corr}}$) inspired by the STOI. The clean target speech and the noisy speech are decomposed per auditory frequency (using a gammatone filterbank), then their envelopes are extracted and passed through a modulation filterbank. The outputs of the latter filterbank are processed by an envelope extraction followed by a logarithmic compression. Then, the correlation between the clean and noisy speech is estimated within each time frame. The long-term broadband correlation is obtained by integrating the correlation values across time and modulation frequency as well as auditory frequency. The correlation is converted to intelligibility score using a mapping function. The model was tested on datasets collected with NH listeners investigating the influence of additive noise only (involving three types of noise), reverberation, noise reduction (using 2 methods and including different types of noise) and phase-jitter distortion (i.e., the signal is multiplied by a time-dependent phase jitter function). The model accurately predicted the conditions with the additive noise (r = 0.97 and $MeanErr$ = 1.9 dB), noise reduction (r = 0.82; $MeanErr$ = 0.6 dB for the first method and r = 0.79; $MeanErr$ = 12.1% for the second method) and phase-jitter distortion (r = 0.97; $MeanErr$ =

19%). The model failed to account for the effect of reverberation but removing the framing of the signals and computing the correlation on the long-term signals figured out this issue ($r = 0.94$; $MeanErr = 1.1$ dB).

Biberger and Ewert (2017) developed the multi-resolution generalized power spectrum model (mr-GPSM) that combines the (S+N)NR in the modulation domain (using the mr-sEPSM) with the (S+N)NR in the auditory frequency domain to predict intelligibility. The model inputs are the noisy speech and the noise alone. The signals are passed through outer and middle ear filters, are decomposed by auditory frequency (using a gammatone filterbank) and their envelopes are extracted. These envelopes are processed by the mr-sEPSM to derive the (S+N)NR in the modulation domain. The (S+N)NR in the auditory frequency domain is computed by calculating the (S+N)NR in each time frame using the power of the signal envelopes, then averaging the values across time and integrating across frequency to get a broadband (S+N)NR. The final broadband (S+N)NR is defined by selecting the higher broadband (S+N)NR resulting from mr-sEPSM or from the auditory frequency domain only. Then, the (S+N)NR is converted into intelligibility threshold using a mapping function. The mr-GPSM was tested on the same dataset as the mr-sEPSM as well as another dataset collected with NH listeners varying the type of spectro-temporal modulation of the noise and the gender combination of target speech and masker. Over the entire datasets, the mr-GPSM predictions were more accurate compared to the mr-sEPSM predictions ($r = 0.94$; $RMSErr = 3.8$ dB vs $r = 0.86$; $RMSErr = 5.2$ dB, respectively).

Relaño-Iborra et al. (2019) combined a model of the auditory periphery modelling the non-linear behavior of the auditory filters (instead of using the linear gammatone filterbank) with the sEPSM$^{\text{corr}}$. This allowed to improve the predictions of the sEPSM$^{\text{corr}}$ on the conditions involving additive noise only ($r = 1$; $MeanErr = 1.2$ dB), phase-jitter distortion ($r = 0.97$; $MeanErr = 6.4\%$) and one of the two methods of noise reduction ($r = 0.6$; $MeanErr = 1.4$ dB for the first method and $r = 0.88$; $MeanErr = 10\%$ for the second method).

II.2.2 Binaural models tested only with normal-hearing listeners

An exhaustive literature review of binaural speech intelligibility models for NH listeners is provided by Lavandier and Best (2020) grouping and distinguishing the models by their rationale to quantify intelligibility (SNR in the frequency domain, SNR in the modulation domain, correlation between the clean and degraded target...). Here, the binaural models are described with the same rigor as the previous ones. However, the detail of the experimental design involved in the model validation is not included to the review for sake of brevity and also to avoid redundancy with the work of Lavandier and Best (2020). This section focuses on the model tested only with NH listeners and conditions without hearing-aid processing.

II.2.2.1 The binaural STI

Van Wijngaarden and Drullman (2008) developed a binaural STI (BSTI). The modulation transfer function is computed in each frequency band (such as in the STI). According to the frequency band, either the modulation transfer function is chosen as the higher modulation transfer function across ears or the modulation transfer function is defined using a correlogram (that can be interpreted as the better-ear listening or binaural unmasking component of the model, respectively). The modulation transfer functions are then integrated across frequency to obtain a binaural STI score. The implementation of this model is simple and fast but it was tested only with NH listeners and the conditions did not involve hearing-aid processing.

II.2.2.2 The short-time equalization cancellation model

Wan et al. (2010) followed by Wan et al. (2014) developed the short-time equalization-cancellation (STEC) model. The model inputs are the target and noise signals that are decomposed into time-frequency units. An EC mechanism is applied to each unit and then the signals are reconstructed across time frame in each frequency band. Then, the highest (long-term) SNR is selected between the binaural-enhanced SNR, the left and the right monaural SNRs. The SNRs are used to compute a SII score that is then transformed into a SRT using a mapping function.

It is worth noting that the mapping function was changed for each type or number of maskers. Also, the dip listening advantage cannot be predicted given that the SNRs are computed using the long-term spectrum of the signals.

II.2.2.3 The binaural speech-based envelope power spectrum model

The binaural version of the model mr-sEPSM (B-sEPSM) was developed by Chabot-Leclerc et al. (2016). The B-sEPSM takes the noisy target signal as input as well as the noise signal alone. The signals are filtered per auditory frequency band (using a gammatone filterbank) and the envelope of each channel output is extracted before being processed by an EC stage to reduce the noise. Afterwards, the modulation spectrum of the signals are computed per time frame and frequency band to derive a (S+N)NR. The higher (S+N)NR between the binaural and monaural channels is considered as output of each time-frequency unit. An average across time and frequency is then calculated and converted to speech intelligibility thresholds using a mapping function.

Nothing is implemented to take into account listener's hearing loss. The monaural model on which the B-sEPSM is based on (Jørgensen et al., 2013) is able to take into account the effects of reverberation on the target as well as spectral noise suppression on intelligibility for NH listeners. However, it is yet to confirm the backward compatibility of the B-sEPSM for these effects.

II.2.2.4 Binaural distortion-weighted glimpse proportion metric

Tang et al. (2016) developed the binaural distortion-weighted glimpse proportion (BiDWGP) metric that evaluates the proportion of target glimpses in the spectro-temporal domain. A glimpse is a time-frequency region in which the target information is clear and available to improve intelligibility (Cooke, 2006). The model inputs are either the binaural target and masker signals or the monaural target and masker signals plus their spatial locations in order to estimate the corresponding binaural signals. The binaural signals are decomposed in the time-frequency domain. A glimpse is available in a spectro-temporal region where the target level is higher than the NH hearing threshold and the target level plus the binaural unmasking advantage (computed using the formula from Culling et al., 2005) exceeds the masker level plus 3 dB. Thereby, binaural unmasking is taken into account in the glimpse definition. Then, the binaural glimpses are computed by taking all the regions presenting a glimpse at the left and/or right ear(s). The binaural glimpses are time-averaged and integrated across frequency using a SII weighting as well as a distortion weighting, which quantifies by how much the masker signal distorts the target signal. The model outputs can be correlated with speech intelligibility measures, however, it is still to be defined how converting the metric results into meaningful intelligibility scores.

II.2.2.5 An EC stage with a glimpsing approach

Mi and Colburn (2016) combined a glimpsing model using an EC stage with the CSII (referred to as G_{EC}+CSII in Table II.a) to predict intelligibility in multitalker mixtures. The model inputs are the binaural noisy target as well as the target ILD and ITD. The signals are decomposed to apply an EC stage in each time-frequency unit using the target ILD and ITD. In other words, the EC stage equalizes and cancels the target signal. The residuals are evaluated and compared to a frequency-dependent threshold to generate a time-frequency matrix that indicates the time-frequency units where the target is enough dominant to be useful for intelligibility (i.e., glimpse). Finally, a noise-reduced target signal is constructed by selecting only the target-dominated regions of the noisy target. The resulting signal is taken as input to the CSII to predict intelligibility. This approach is interesting because it assumes that the listener need only to know the target position, which is likely in real life, to define its ILD and ITD to be able to segregate the speech from the competing sources.

II.2.2.6 Periodicity-based glimpsing model

Josupeit and Hohmann (2017) developed a model (referred to as PerG in Table II.a) to predict word recognition, speech localization and talker identification in a multi-talker environment.

Their rationale to solve such tasks was the importance of periodicity in the auditory scene analysis. The model analyzes the target+masker signals using an auditory processing stage to estimate the periodicity of the signals. A glimpse is available when the periodicity is high enough. The robust glimpses are compared to templates that are generated with the corpus word in quiet at different locations. Thereby, they can identify the talker (that utters a particular call sign), the location and the target words. The approach of this model is different compared to the previous binaural models. First, It does not consider an EC stage to improve target intelligibility. Second, a glimpse here is defined using four periodicity-based features (rather than being SNR-based) that are related to the pitch, spectral shape, ILD and ITD of the periodic sound.

II.2.3 Binaural models for hearing-impaired listeners or hearing-aid processing

The binaural models described in this section are those that consider the effects of hearing impairment and/or hearing-aid processing. The different versions of a given model are described even if they were not tested on data collected with HI listeners or involving hearing-aid processing. However, for most of these model versions, further details can be found in Lavandier and Best (2020).

II.2.3.1 The short-time binaural speech intelligibility model

This binaural speech intelligibility model (BSIM) has been originally developed by Beutelmann and Brand (2006) but its implementation was revised by Beutelmann et al. (2010), who provide a short-time version of the BSIM (stBSIM) that considers the fluctuations in the masker envelope. The target and noise signals at the listener's ears as well as the listener's audiogram are considered as model inputs. The external signals are decomposed in the time-frequency domain (using a gammatone filterbank and half-overlapping Hann windows). An internal noise is designed based on the spectral shape of the audiogram plus a gain parameter to define its broadband absolute level. Then, it is added to the external noise signal to take into account the hearing loss of the listener. The signals are processed by a EC stage, a delay and a gain are applied to the signals at the left and right ears (equalization stage) and then they are subtracted (cancellation stage). The formula defining the binaural SNR is designed in a way that it tends to the monaural SNRs in the case that the optimum gain and delay tend to $\pm\infty$. The binaurally enhanced SNRs are used in the SII calculation. The SII scores are then converted into speech intelligibility scores using a mapping function.

The model was tested in the original publication on a dataset measuring SRTs for NH and HI listeners in four different simulated acoustic rooms (varying the reverberation). Three types of noise and three spatial configurations were tested. The model was able to predict the trends across groups and conditions ($r = 0.88$), but the predicted SRTs were on average 3.4 dB lower than the data, and predictions were generally less accurate for HI listeners. The model also successfully predicted the influence of 4 beamformers on speech intelligibility in diffuse noise for NH listeners ($r = 0.9$; $RMSErr = 0.8$ dB Hauth et al., 2018). The model is yet to be tested on datasets involving HI listeners tested under conditions with non-linear hearing-aid processing.

Rennies et al. (2011) extended the BSIM (long-term version, Beutelmann et al., 2010) to account for the effect of target smearing due to room reverberation (see Sec. II.2.1.1). To do so, they investigated three ways of combining the BSIM with different monaural indices known to predict the effect of target smearing. This model is referred to as combinedBSIM in Table II.a. These indices are (1) the modulation transfer function that quantifies how well the temporal modulations of a speech are preserved during the transmission; (2) the definition that is the ratio between the energies of the early reflections and of all the reflections; and (3) the useful-to-detrimental ratio that considers the early reflections as useful for intelligibility while the late ones are not, and can be considered as another masker. The limit L between early and late reflections was a free parameter in the study and took the values of 50, 80 and 100 ms. The modifications of the BSIM to include the modulation transfer function and the definition are similar: both indices degrade the binaurally-enhanced SNRs resulting from the BSIM and then the SII is computed. The useful-to-detrimental ratio is used to generate an useful target signal that input to the BSIM and the detrimental speech that is added to the noise to form

the second input to the BSIM. Note that the three modified models require having the room impulse responses as input in addition to the target and masker signals. Even though the scope of application is extended with this combination model (compared to the BSIM), it has not been tested on datasets involving non-linear hearing-aid processes nor HI listeners.

The BSIM2020 combines the long-term BSIM (but using the implementation of Beutelmann and Brand, 2006) with the SRMR in order to provide a non-intrusive binaural processing model (Hauth et al., 2020). The model input is the binaural target+masker mixture. The signals are band-pass filtered then processed by an EC stage below 1,500 Hz, the EC parameters (gain and delay) are estimated to equalize the signals and then the cancellation stage is either done by adding or subtracting both channels. The operation that maximizes the SRMR is selected. Above 1,500 Hz, the ear leading to the best SRMR is chosen and in case SRMRs are similar at both ears, the ear leading to the lower level is selected. Then, the EC channels and the better-ear channels are combined to reconstruct a monaural signal. Hauth et al. (2020) decided to use the SII as back-end intelligibility model, which requires the monaural target and masker signals separately. Hence, at this stage the model is intrusive but the monaural binaurally-enhanced target and masker signals — at the input of the SII — are generated using the EC parameters that were defined by the blind binaural processing. That is why in Table II.a, the inputs are "S+N then S & N" and the approaches are "SNR_{MD}then SNR_{AFD}". The SII scores are converted into SRTs using a mapping function. The model was tested on two datasets but neither HI listeners nor non-linear hearing-aid processing were considered.

II.2.3.2 The better-ear SNR enhanced by binaural unmasking advantage model

Lavandier and Culling (2010) developed a model that adds the binaural unmasking advantage to the better-ear SNR to predict intelligibility. The target speech and noise signals are the input to the model, which are decomposed per frequency band (using a gammatone filterbank). Then the SNR at the better ear (i.e., the higher SNR between both ears) is computed as well as the binaural unmasking advantage (using the binaural unmasking level difference (BMLD) formula from Culling et al., 2005) which is added to the better-ear SNR. After integration across frequency using a SII weighting, the model output can be compared to intelligibility thresholds using a mapping method.

The model of Lavandier and Culling (2010) was further developed by Collin and Lavandier (2013, referred to as st(BE+BU) in Table II.a) applying the computation of the model within short-time frames to predict the influence of the modulation in the masker envelope. The model was used by Cubick et al. (2018) to predict the effect of spatial distortion when listening through hearing aids in NH listeners. The trends in the data were well predicted ($MeanErr$ = 0.6 dB; $MaxErr$ = 1.1 dB). However, the involved conditions investigated only the influence of the hearing-aid microphone without any hearing-aid processing. This model version is not able to account for the effects related to hearing loss such as reduced audibility.

Cosentino et al. (2014) proposed a binaural version of the SRMR (referred to as B-SRMR in Table II.a) combining the SRMR with the model developed by Lavandier and Culling (2010). The better-ear component is computed selecting the higher SRMR across ears in each frequency band. The binaural unmasking advantage is calculated using the BMLD formula from Culling et al. (2005). They assumed that the target IPD is equal to 0. The masker IPD and interaural coherence are defined by first computing the running IPD and interaural coherence time functions, then preserving the time samples that lead to high coherence (coherence-based filtering) to finally estimate the interaural cues considering the values that occur the most in the running function. The better-ear listening component of this model is based on the SRMR that estimates a SNR in the modulation domain (SNR_{MD}) while the binaural unmasking component computes a BMLD in the auditory frequency domain, which can be seen as a difference in SNR_{AFD}. That is why the approach in Table II.a is referred to as "SNR_{MD}+ SNR_{AFD}". Despite the monaural SRMR predicted a dataset collected with HI listeners, the binaural version was tested only with NH listeners and does not consider listener's hearing threshold.

Leclère et al. (2015) proposed a model (referred to as combined(BE+BU) in Table II.a) to account for the effect of temporal smearing and binaural de-reverberation (in presence of a reverberant target, intelligibility is improved with binaural listening compared to diotic listening, Lavandier and Culling, 2008). The model is based on the model of Lavandier and Culling (2010)

with the implementation of Lavandier et al. (2012) that uses the binaural room impulse responses (BRIRs) at the target and masker locations. It combines the existing model with the concept of useful early reflections and late detrimental reflections for speech intelligibility. Only the part of the speech signal resulting from the early reflections is considered as target speech and the masking noise is composed of the masking sources and the late reflections of the speech signal.

Lavandier et al. (2018) modified the model for NH listeners (Collin and Lavandier, 2013, with the implementation used in Cubick et al., 2018) to take into account audibility (referred to as st(BU+BE)$_{\text{Aud}}$ in Table II.a). In addition to the target and masker signals, the listener's audiogram is now an input to the model to create an internal noise level at each ear. The internal noise is spectrally shaped using the listener's audiogram and a fitting gain parameter is added to define the overall internal noise level. Based on the incoming signals and the internal noise levels, the better-ear SNR is computed per time frame and frequency band using the higher level between external and internal noises. This better-ear SNR is then further enhanced by adding the binaural unmasking advantage, but only if the target and external noise levels are above the internal noise level at both ears. The results are averaged along time and integrated across frequency using a SII weighting. The same mapping method as the previous version is used to compare the intelligibility model outputs to measured intelligibility thresholds.

This model was tested on datasets highlighting the effect of varying audibility on speech intelligibility for HI and NH listeners in the presence of stationary and envelope-modulated noises in different spatial configurations. The resulting performance statistics showed that the model accurately predicted SRTs with r greater than 0.91, $MeanErr$ lower than 1.7 dB and $MaxErr$ equal to 2.6 dB across experiments. The drawbacks of this model are addressed in one chapter of the PhD, hence, they are presented latter in Sec. II.3.

II.2.3.3 The modified binaural short-time objective intelligibility model

The deterministic binaural short-time objective intelligibility (DBSTOI) model, which is the binaural version of the STOI, bases its speech intelligibility prediction on a distortion measure of the speech envelope (Andersen et al., 2016). This model applies a short-time frequency analysis on the noisy target and the clean target signals at the two ears. These short-term spectra are then processed by an EC stage, resulting in enhanced monaural signals, and the envelopes of the outputs are extracted. Finally, the correlation between the envelope of the noisy target and the clean target is estimated and then averaged across time and frequency to obtain a time-averaged broadband correlation coefficient. The EC parameters are chosen in a way to maximize the time-averaged broadband correlation. The correlation is mapped to an intelligibility score to be compared to a speech intelligibility measure.

Andersen et al. (2018) showed that the model tended to overestimate intelligibility when the sources were spatially distributed and presented at low SNRs. They analytically demonstrated that the way to choose the EC parameters induced an unwanted bias so they changed the strategy to remove this bias. The modified DBSTOI (MBSTOI) improved the predictions.

The model validation involved five datasets measuring percentage of correct words for NH listeners using 262 conditions that varied the SNR, the spatial locations of the sources, the type of signal processing (simulating spectral noise suppression or beamforming as used in hearing aids) and the masker type. They showed that the model can predict the effect of the tested factors with $RMSErr$ that varied between 6.3% to 7.7% depending on the dataset. They also successfully predicted SRM ($MeanErr$ = 0.8 dB[2]) for frontal speech masked by a SSN presented at 10 different azimuths including 0°. In comparison, Andersen et al. (2018) applied the B-sEPSM and obtained a $MeanErr$ of about 4.9 dB. Andersen et al. (2018) suggested to further test the MBSTOI on conditions with reverberation and fluctuating interferers. Also, even though the MBSTOI is able to take into account the effect of some hearing-aid processes, nothing is implemented to account for the effect of hearing loss.

II.2.3.4 The simulation framework for auditory discrimination experiments

The simulation framework for auditory discrimination experiments (FADE) applied to speech intelligibility experiment in binaural listening was developed by Schädler et al. (2018). This

[2]Value calculated for the current manuscript.

model considers a different approach compared to the other binaural models in the current review because it uses an automatic speech recognition algorithm to predict speech intelligibility thresholds and it does not consider an EC stage to internally improve target intelligibility. However, the scope of application is interesting, and worth to be mentioned.

The automatic speech recognizer needs a training phase in which it learns the spectro-temporal pattern of each tested word at each considered SNR for each experimental condition (e.g., hearing-aid setting or noise type). The spectro-temporal representations of the word+noise mixture are degraded by the individual audiogram removing the information that are not audible. To model binaural hearing, the features of the spectro-temporal representations at each ear are kept by the automatic speech recognition algorithm (this can be interpreted as a better-ear listening mechanism). At this stage, the training phase is over and the prediction phase starts. The sentences played during the experiment are analyzed by the automatic speech recognizer and then the percentage of correct words as a function of SNR can be derived.

The model was tested on a dataset involving NH and HI listeners, three masker types (20-talker babble noise, cafeteria ambient noise recording and speech masker) and eight types of hearing-aid processing (including non-linear processes). The target speech was always simulated in front of a listener. Non-linear amplification was applied to the stimuli for the HI listeners to compensate partly for their hearing loss. The FADE was more accurate to predict the effect of hearing-aid processing for NH listeners ($r = 0.93$; $RMSErr = 1.9$ dB for the 20-talker babble noise and $r = 0.96$; $RMSErr = 2.9$ dB for the cafeteria ambient noise) than for HI listeners ($r = 0.79$; $RMSErr = 2.1$ dB for the 20-talker babble noise and $r = 0.91$; $RMSErr = 2.8$ dB for the cafeteria ambient noise). The model was not able to predict accurately the conditions with the speech masker because it could not differentiate target speech from the speech masker. When considering the individual SRTs of HI listeners (rather than averaging SRTs across listeners), the model failed to predict the influence of hearing-aid processing ($r = 0.40$; $RMSErr = 3.2$ dB for the 20-talker babble noise and $r = 0.32$; $RMSErr = 3.9$ dB for the cafeteria ambient noise).

The strengths of this model are that it is not intrusive and the output does not require any mapping method to be transformed into meaningful intelligibility thresholds (it is straight a percentage of correct identified words). It is also able to predict the effect of hearing-aid processing in a cafeteria ambient noise that is an essential achievement for practical use. However, the model needs further development to predict individual intelligibility in real-life scenario. The main drawback of this model is the loss of an understanding of the involved perceptual mechanism by using an automatic speech recognition algorithm.

II.2.4 Summary and outlook

Figure II-8 purposes to summarize the model presented in the current review using a visual representation. It is a double-entry table, the rows list the model approaches to predict intelligibility while the columns highlight their ability to predict monaural or binaural listening. The columns also highlight if the model was tested on conditions involving non-linear hearing aid processing and/or if it accounts for the effect of hearing impairment on intelligibility. Compared to Table II.a, this figure shows using arrows the relation between models from the same family, especially, a double arrow heads toward the binaural version of a monaural model.

The diversity of approaches to predict intelligibility highlights that the auditory system is a complex system to model and its behavior is likely a combination of these approaches as suggested by the hybrid models from Biberger and Ewert (2017); Hauth et al. (2020); Cosentino et al. (2014), which consider the SNR in the frequency domain as well as in the modulation domain. In this way, every model makes its contribution to understand how the auditory system could work in a certain acoustic set-up to understand speech. From a point of view of model users, the plurality of models and required inputs are handy because they can choose which one is more relevant to use to predict intelligibility in a given context. That is why it is also important to highlight the advantages and limitations of each model. Obviously, a single model that requires only target+noise mixture to predict speech intelligibility in any scenario would be better because the user would not have to choose which model to apply, however, this model does not exist yet and the models presented in the current review are likely helping to understand how to design such a model.

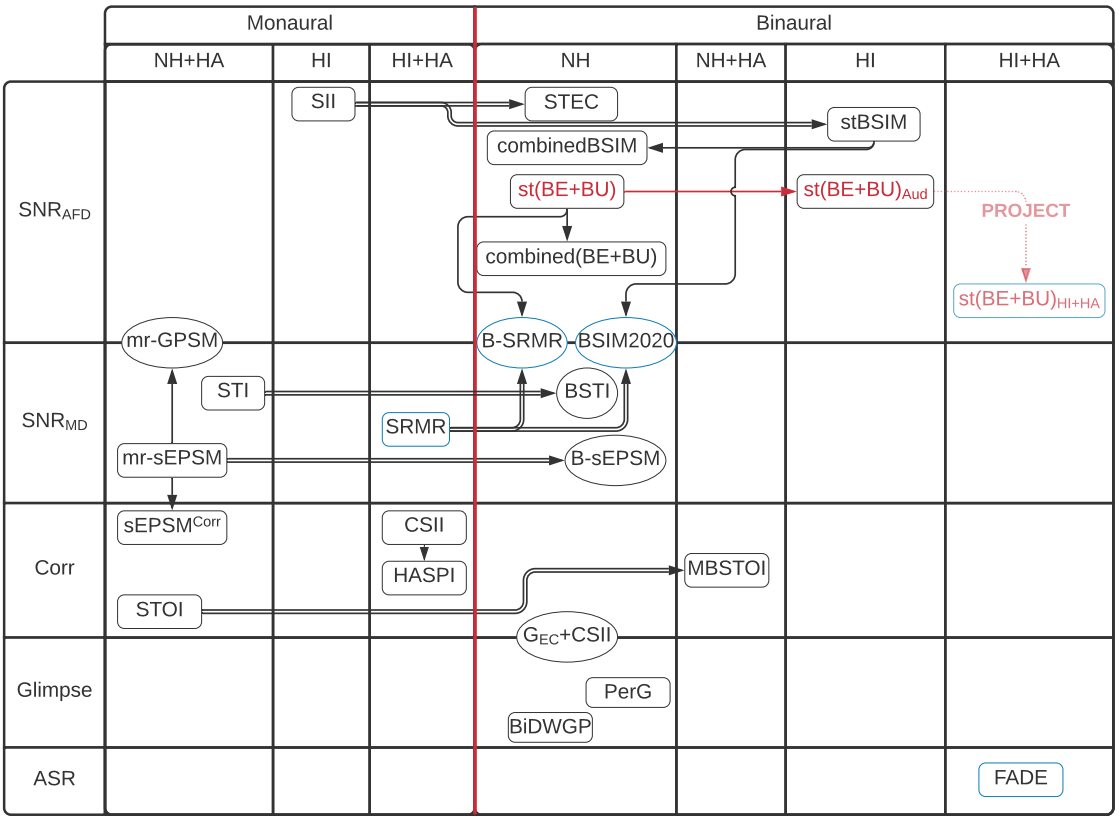


FIGURE II-8: Visual representation of the models presented in the current review as a function of the approach used to predict intelligibility as well as the listening scenarios in which they were tested (listening conditions and listener hearing profile). The hybrid models are placed over the two approaches considered to predict intelligibility. The non-intrusive models that require only the target+noise mixture as input are distinguished with blue boxes. The arrows between models highlight the fact they belong to the same family, and the double arrows are used to additionally show the binaural versions of the monaural models. The models used and further developed during the PhD are shown using a red font. The st(BE+BU)$_{\text{HI+HA}}$ (in faded red) is the long-term goal of the project, which is developing a model that is able to predict speech intelligibility in noise for HI listeners wearing hearing-aid devices.

Binaural speech intelligibility models play an important role in understanding auditory scene analysis. However, the models still need to be developed to predict real-life scenarios for any kind of listeners. Throughout the review presented above, only few models considered some basic effects of hearing impairment (Beutelmann et al., 2010; Schädler et al., 2018; Lavandier et al., 2018) or were tested on datasets involving hearing-aid processes (Schädler et al., 2018; Andersen et al., 2018; Chabot-Leclerc et al., 2016). Moreover, only few models attempted to predict speech-on-speech masking (Wan et al., 2014; Mi and Colburn, 2016) or take as inputs solely the target+noise mixture (highlighted with a blue box in Fig. II-8), which reflects the actual signals the ears receive in real life. The model approach from Mi and Colburn (2016) or Hauth et al. (2020) suggests that the auditory system can segregate the target signal from the target+masker mixture using an EC mechanism to be able then to analyze the signal alone.

The majority of the models considers only bottom-up processes (related to signal and sensory paths) but not top-down processes (related to the use of knowledge acquired by experience, such as vocabulary). Somehow, the models developed by Josupeit and Hohmann (2017) and Schädler et al. (2018) implement a top-down process because both involve the use of spectro-temporal representations of word in noise, which need to be defined before predictions. Furthermore,

other factors are not taken into account in the above models such as the effect of aging, attention, listening effort, visual cues... The future models would need to take these effects into consideration in order to predict intelligibility in a cocktail-party environment for any kind of listeners.

II.3 The role of this PhD project

The main goal of this PhD project is to better understand how the different aspects of hearing loss affect the perceptual mechanisms related to speech intelligibility in noise. To address this goal, a speech intelligibility model is developed and its predictions are compared to data measured in a range of critical conditions with NH and HI listeners. Thereby, existing datasets are used as well as new data, in particular to better understand the contribution of binaural unmasking to speech intelligibility for HI listeners.

The PhD project focuses on the models from Collin and Lavandier (2013) and from Lavandier et al. (2018, an overview of the models is provided in Fig. II-9) and addressed a number of limitations. The first model was validated in the original publication (Collin and Lavandier, 2013) using 3 datasets measured in NH listeners, varying the type of listening (diotic or dichotic), the spatial separation between the sources, the target azimuth and the modulation depth in the noise envelope. The model accurately predicted the measured SRTs in each dataset, with a correlation r greater than 0.85, $MeanErr$ lower than 0.7 dB and $MaxErr$ lower than 1.6 dB across experiments. However, the parameters of the model were not defined thoroughly and this needed to be taken into consideration before developing a model for HI listeners.

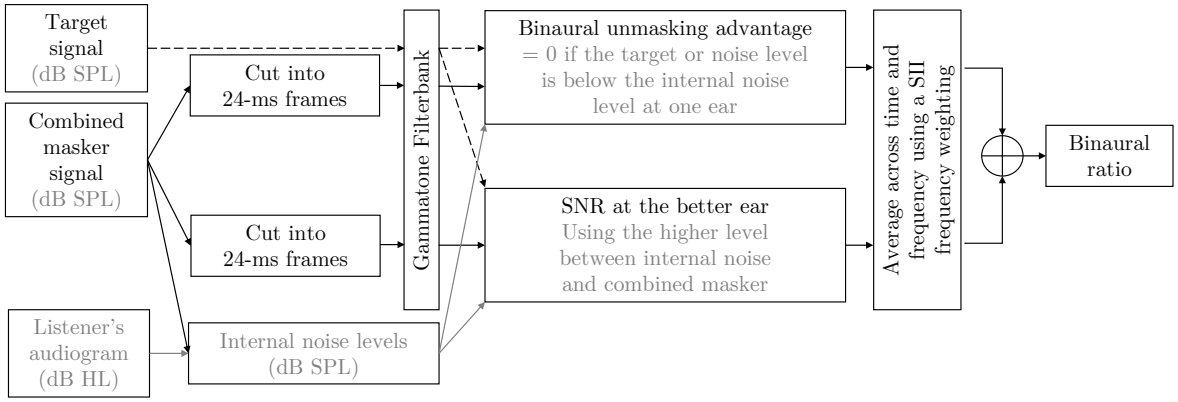


Figure II-9: Block diagram of the models that have been considered in the PhD project. The black font represents the model designed only for NH listeners Collin and Lavandier (2013) and the modifications associated with the extension to HI listeners are highlighted in grey Lavandier et al. (2018).

Regarding the limitations of the model from Lavandier et al. (2018), first, the gain parameter defining the absolute level of the internal noise that is used to simulate a listener's hearing loss was different for NH and HI listeners to obtain good predictions for both groups. This effectively resulted in two different models, one for NH and one for HI listeners so that the difference between listener groups could not be predicted. Moreover, this model version was validated only on datasets involving ILD but no ITD. Hence, the binaural unmasking component of the model was not tested yet. These datasets were collected with anechoic stimuli, meaning that the model had yet to be tested on more realistic stimuli that include room reverberation, as well as non-linear hearing-aid processes (as shown in faded red in Fig. II-8).

The following chapters describe the research that has been conducted during this PhD. Chapter III presents a further validation of the model for NH listeners developed by Collin and Lavandier (2013), including an optimization stage of the model parameters as well as the consideration of binaural sluggishness. This should be seen as a deep introduction of the original model that is further developed in the following chapters to take into account hearing loss. Chapter IV describes the modification of the model from Lavandier et al. (2018) to provide a single model

version that predicts speech intelligibility in modulated noises for NH and HI listeners. Chapter V is about our investigation of the contribution of binaural unmasking to speech intelligibility in noise for NH and HI listeners. This includes the data collected at Macquarie University, the validation of the binaural unmasking component of the model as well as a modelling suggestion to take into account the effect of presentation level on binaural unmasking for NH listeners. Chapter VI is a general discussion about the PhD achievements and outlook.

III

Further validation of a binaural model predicting speech intelligibility against envelope-modulated noises

The study purposed to thoroughly test the robustness of the model developed by Collin and Lavandier (2013) to predict speech intelligibility amongst noise in rooms using five experiments. The influence of the parameters involved in the model was evaluated varying their values. This also allowed to find the set of parameter values that led to the best fit between data and predictions. Then, the revised model was verified using an experiment from the literature including 40 experimental conditions. I started this study during my MSc research project but most of the work presented in this chapter was done at the beginning of my PhD. The content of this chapter is published in the journal Hearing Research (Vicente and Lavandier, 2020).

III.1 Introduction

In the presence of multiple envelope-modulated noises, the SNR at each ear can change quickly over the time, so that the better ear is not always the same. As mentioned in Sec. II.1.3.2, better-ear glimpsing denotes the ability of the auditory system to benefit of these variations, switching to the better ear. The exact nature of this mechanism is not yet clear. Better-ear glimpsing could be a "true" binaural mechanism in which the auditory system compares the SNR at the two ears and then switch back and forth from one ear to the other to follow the ear with the best SNR (Culling and Mansell, 2013). It could also result from two simultaneous monaural mechanisms at each ear, providing the SNRs at both ears (Brungart and Iyer, 2012). These two interpretations might not involve the same time constants or limitations in terms of following changes across time at the ears. The binaural system appears to be sluggish compared to the monaural system (Grantham, 1982). This binaural "sluggishness" corresponding to a poorer temporal resolution can be modeled by using a longer time window when describing the mechanism. Values ranging from 40 to 250 ms have been proposed for a binaural temporal window (Culling and Summerfield, 1998; Culling and Mansell, 2013; Grantham and Wightman, 1979; Hauth and Brand, 2018). In contrast, the time constant usually used to describe the monaural system is about 8-13 ms (Moore et al., 1988; Plack and Moore, 1990). Culling and Mansell (2013) provided evidence that better-ear glimpsing could be "truly" binaural and rely on switching across ears, since they found that performance was highly dependent on the required switching rate, and that this binaural switching could be rather sluggish.

The current study concentrated on the model proposed by Collin and Lavandier (2013) to predict speech intelligibility against multiple envelope-modulated noises in rooms. This model has four parameters: the size of the temporal windows used for computing the better-ear listening and binaural unmasking components, the degree of sampling of the spectral information, and a SNR ceiling used when estimating better-ear listening (see Sec. III.3 for the description of the model). The sizes of the temporal windows can be seen as the temporal resolution for computing the better-ear SNR and the binaural unmasking advantage. The spectral sampling is the number of filters the model considers per ERB unit. This sets the amount of the spectral information of the signals. The SNR ceiling sets the highest SNR allowed when computing the better-ear SNR, it is likely related to the speech material and the tested modulation depth. The influence of these parameters has not been thoroughly evaluated yet, only a few values of SNR ceiling were tested while the other parameters were not varied. Moreover, the model has only been evaluated in a limited number of conditions, all involving the same speech material.

The main aim of the present study was to test the robustness of the model proposed by Collin

and Lavandier (2013), considering critical conditions and also different speech materials (see Sec. III.2.1). The influence of the model parameters was evaluated using an approach inspired by a variance-based sensitivity analysis (see Sec. III.4.1). It involved the predictions — varying the model parameters — of four previously published experiments and one specifically designed for the present study. The results allowed highlighting the potential interactions between model parameters, as well as the parameter values leading to the best predictions across the five experiments. Another aim of the study was to analyze in details the model predictions, thus highlighting the effects and configurations accurately predicted and the remaining limitations of the model. The size of the temporal window used to model binaural unmasking was revised, so that binaural sluggishness could be partly described. This study also tried to play its part in discussing the controversial concepts of better-ear glimpsing mentioned above. Finally, the optimized model was tested using an additional dataset not used to define its parameter values (see Sec. III.5).

With the proposed model, we want to provide a metric able to predict speech intelligibility in real-life listening. This is why we considered conditions involving running speech for the target, speech modulations for the maskers and real-room reverberation. However, it is sometimes useful to consider unrealistic synthetic stimuli. Isolating better-ear listening and binaural unmasking is not realistic, but tests whether the model can predict both effects correctly.

The proposed model is made available to the community and a code can be downloaded here: https://mathieulavandier.wordpress.com/home/models/.

III.2 Data

III.2.1 Data sets used to test the model parameter

Five experiments were used to test the model parameters. The experiments 1 and 2 of Culling and Mansell (2013) are abbreviated CM1 and CM2, the experiments 1 and 4 of Collin and Lavandier (2013) are CL1 and CL4, the experiment run in the present study is VL. A summary of the design of each experiment is presented in Table III.a, for more details referred to the related publications (Culling and Mansell, 2013; Collin and Lavandier, 2013; and Appendix A, respectively). The "co-located" condition will refer to the configuration where target and noise(s) are at the same spatial position, otherwise the configuration will be referred to as "separated". Positive azimuths correspond to the right side of the listeners. All the noises used as masking sources were SSNs.

Two experiments from Culling and Mansell (2013), i.e. CM1 and CM2, were chosen in order to test the model in anechoic conditions and in presence of artificially modulated maskers. In particular, CM2 investigated the influence of binaural sluggishness on better-ear listening and binaural unmasking independently, which is relevant to test the temporal resolutions used in the model. The SRTs are displayed on Fig. III-2 as a function of noise azimuth, number of noises and type of noise modulation for CM1 and on Fig. III-3 as a function of modulation rate for CM2.

CL1 was chosen to test the model performance at predicting the effect of reverberation on speech intelligibility in the presence of modulated noise and to consider envelope modulations more characteristic of real speech (rather than artificial modulations; see Table III.a). The measured SRTs are plotted as a function of masker distance in Fig. III-4, each panel corresponds to a modulation depth for the noise. CL4 was considered because it involved reverberation and speech modulations for the noises, but also asymmetrical configurations in which binaural hearing and SRM were involved. Figure III-5 presents the SRTs measured for each type of masking noise.

VL was designed to evaluate the model at predicting the influence of reverberation filling in the masker modulation gaps in an asymmetrical condition (see Appendix A[3]). The better-ear component of the model was tested on its own and in combination with the binaural unmasking component. The measured SRTs are plotted in Fig. III-6 as a function of the noise position, with one panel for each type of noise modulation.

[3]The experimental methods of this experiment are available in Appendix A to allow focusing on the modelling, which is the core of the study.

Exp.	Number of Noises	Noise modulation	Noise Distance in room	Noise Azimuth	Cues available
CM1	1 or 2	Steady-state or modulated (10-Hz square wave, 50% duty cycle)	Anechoic	$0°^T$, 105° or ± 105°	ITD+ILD
CM2	1 or 2	Steady-state (0 Hz) or modulated (1-, 2-, 5-, 10-, or 20-Hz square wave, 50% duty cycle)	Anechoic	$0°^T$ or ± 105°	ITD+ILD, ILD-only or ITD-only
CL1	1	Steady-state or modulated (broadband envelope of 1, 2 or 4 voices)	0.65^T, 1.25 or 5 m	$0°^T$	ITD+ILD
CL4	1 or 2	Steady-state or modulated (broadband envelope of 1 or 2 voices)	0.65^T m	$0°^T$, 25° or ± 25°	ITD+ILD
VL	1	Steady-state or modulated (broadband envelope of 1 voice)	0.65^T or 5 m	$25°^T$ or -25°	ITD+ILD or no ITD/no tail

Table III.a: Summary of the experimental designs used to test the model parameters. The superscript 'T' indicates the target's distance and azimuth and defines the co-located condition. The last column indicates the nature of the binaural cues available in the tested signals.

III.2.2 Data set used to validate the revised model

In order to validate the revised model, the experiment of Ewert et al. (2017) was considered. The target was always simulated in front of the listener at 0.8 m. Two maskers were involved, either co-located with the target or symmetrically placed on both sides of the listener at ± 60°. Six types of masker were tested, but only the four energetic maskers are considered here. Our model is not designed to predict the effects of informational masking. A steady-state noise and three types of envelope-modulated noise were tested. The modulated noises were generated using: a 8-Hz sinusoidal amplitude modulation, the broadband envelope of a speech signal, and speech modulations incoherent across-frequency (named here sinusoidal noise, 1-voice noise and 1-voice Freq. Inc. noise, respectively). The last type of modulation was obtained by modulating different spectral regions of the noise with different speech envelopes.

Five head-related impulse response (HRIR) conditions were tested. (i) A natural HRIR (ITD+ILD) condition without processing (ii) An ILD-only condition (iii) An ITD-only condition for which the HRIRs spectra at 0° and 60° ("Magnitude 0" and "Magnitude 60", respectively) were averaged across ears (iv) An "Independent" condition was created using the natural HRIR at 0°. One noise source was convolved only with the right ear HRIR while the other was convolved only with the left ear HRIR, resulting in a listening without crosstalk and coherence between ears, thus creating an infinite ILD. (v) Two IMBM (Brungart and Iyer, 2012, or see Sec. II.1.3.2) conditions were also created using the natural HRIR and the independent HRIR (resulting in an IMBM condition or an independent IMBM condition, respectively). The 4 noise modulation types, 5 HRIR conditions and 2 spatial configurations resulted in 40 conditions.

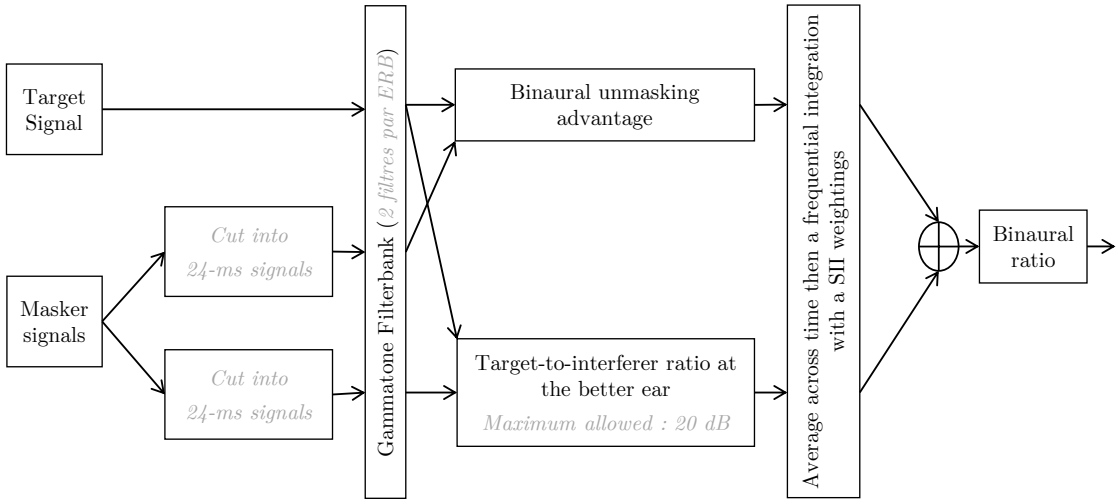


FIGURE III-1: Block diagram of the original model (Collin and Lavandier, 2013), with the parameters tested in the present study highlighted in grey.

The measured SRTs are plotted in Fig. III-7 as a function of the masker type, each panel corresponding to a given type of HRIRs.

III.3 Model description

A block diagram of the model is provided on Fig. III-1. The model takes as inputs the target and combined masker signal at the ears. It predicts the target intelligibility taking into account binaural unmasking and better-ear listening as proposed by Lavandier and Culling (2010). The model computes, per time frame (Rhebergen and Versfeld, 2005) and frequency band, the SNR at the better ear and the binaural unmasking advantage which is added to the better-ear SNR (Collin and Lavandier, 2013). After integration across frequency and averaging across time frames, the model output is a SNR in the corresponding condition, referred to as "binaural ratio" in the following. Differences in binaural ratios can be directly compared to differences in intelligibility thresholds measured in dB. Binaural ratios are first inverted to be compared to SRTs, so that the inverted ratio decreases with the SRT when intelligibility increases. Because only relative differences across conditions can be predicted by the model, a reference needs to be chosen to compare the inverted ratios to the SRTs. A single constant is added to all inverted ratios, so that their mean equals this reference. For each experiment presented here, this reference is the average measured SRT across conditions (Jelfs et al., 2011; Lavandier et al., 2012).

Peaks in the masker signal induce an increase of target masking whereas pauses induce a decrease of this masking. Therefore, the model considers masking energy as a function of time. In order to consider the pauses/envelope modulations in the target speech as important information for its intelligibility (e.g. the gaps between words), the model considers the average level of the target across time rather than its instantaneous level within short-time frames (Rhebergen and Versfeld, 2005). Like Cubick et al. (2018), instead of replacing the target speech by a stationary signal with a similar long-term spectrum and interaural parameters and applying the short-term analysis on this signal (Collin and Lavandier, 2013), the present implementation of the model computes the long-term statistics of the target only once and combines these statistics with the short-term spectrum and interaural parameters of the noise to compute the better-ear and binaural unmasking components within each time frame. Thus, as a model input, the target

sentences at the ears are replaced by an averaged target signal generated by adding at least[4] 60 target sentences (truncated to the duration of the shortest sentence), and this averaged signal is not submitted to the temporal decomposition into short-time frames used for the masker.

The masker signals are cut into frames using half-overlapping Hann windows, before being passed through a Gammatone filterbank (Patterson et al., 1987) with two filters per equivalent rectangular bandwidth (ERB ; Moore and Glasberg, 1983). The bandwidth of the Gammatone filters is about 1 ERB, thus the filters are half-overlapping. Within each time frame and frequency band, the two components of spatial unmasking are modeled, (1) the binaural unmasking advantage is estimated using a formula proposed by Culling et al. (2005), which depends on the masker interaural coherence and on the target and masker interaural phase differences. The target and masker signals are both cross-correlated to derive these interaural parameters. The coherence is taken as the maximum of the cross-correlation function, and the phase difference is obtained by multiplying the corresponding delay by the center frequency of the band. The search of maximum delay in the cross-correlation functions is limited to the range plus/minus half the period of the channel center frequency, so that the model does not predict any binaural unmasking advantage at high frequency (Durlach, 1972). The binaural unmasking advantage is set to zero if the masking noise power is zero at one of the ears in the considered band and frame. (2) The SNR is also computed at each ear, and the best SNR across ear is selected (thus independently for each frequency band and each time frame). A ceiling parameter corresponding to the maximum better-ear ratio allowed by frequency band and time frame is introduced at this stage, to avoid the SNR ratio tending to infinity in masker pauses. Conceptually, this parameter is implemented to explain the fact that a listener does not need an infinite SNR to fully understand the target. (3) The better-ear ratios and binaural unmasking advantages estimated per frequency bands and time frames are then integrated across frequency using the SII weighting (ANSI S3.5, 1997) and averaged across time. Finally, the two values are added to get the binaural ratio.

The first aim of the present study was to test the four parameters of the model (see Fig. III.3): the duration of the Hann window used for computing the binaural unmasking advantage ("BU" in ms), the duration of the Hann window used for computing the better-ear SNR ("BE" in ms) — those are the two temporal resolutions of the model — the number of gammatone filters per ERB (the model spectral sampling "SpecSamp") and finally the ceiling parameter ("Ceiling" in dB). The parameter values were previously set to 24 ms, 24 ms, 2 filters per ERB and 20 dB, respectively (Collin and Lavandier, 2013; Cubick et al., 2018). In particular, the same temporal resolution was used to model better-ear listening and binaural unmasking; whereas the temporal resolution of the two mechanisms (and their susceptibility to binaural sluggishness) was investigated independently here.

III.4 Revision of the model

III.4.1 A method inspired by a sensitivity analysis

One of the aims of the present study was to quantify the relative influence of each parameter of the tested model and to identify potential interactions between these parameters. The method used was inspired by a variance-based sensitivity analysis, which has been described in details by Saltelli et al. (2010). Conceptually, the method consists in computing model predictions while varying the value of its parameters. Then, sensitivity indices can estimate the rate of model output variance due to a given parameter or to an interaction between parameters. For instance, the first order sensitivity index evaluates the direct impact of varying a given parameter on the model output, a second order sensitivity index evaluates the amount of variance in the model output that can be attributed to an interaction between two parameters. The indices are computed so that they all take values between 0 and 1, the sum across all indices is equal to 1 and the higher the index the stronger the influence of the corresponding parameter or interaction. This analysis allows to determine whether strong interactions between model parameters prevent

[4]For CM1 and CM2, 80 and 160 target sentences were used, respectively. Regarding CL1 and CL4, 60 target sentences were averaged. To model the experiment of the present study, 120 sentences of each target type were used. To model the experiment of Ewert et al. (2017), the 120 target sentences of the Oldenburg Satztest corpus were used.

from defining these parameters values independently one from the other, and to identify the most influential parameters of the model.

Five values were tested for each of the 4 model parameters, resulting in 625 combinations. The equivalent rectangular window duration of a Hann window is only half of its full length (Beutelmann et al., 2010). The durations of the Hann windows were here converted into equivalent rectangular duration (ERD). The durations tested were (for both BU and BE): 8, 12, 40, 100, 200 ms (ERD). These values span the range of the monaural and binaural time constants proposed in the literature and mentioned in the Introduction. The values tested for SpecSamp were: 2, 1, 2/3, 1/2, 2/5 filter(s) per ERB. The values tested for Ceiling were 8, 12, 16, 20, 24 dB. Some predictions were done before in order to define the step sizes of SpecSamp and Ceiling. The parameters were varied one at a time using a small step size to roughly evaluate their influences. The final tested values were chosen because they were a good compromise between having a broad range and time-consuming modelling.

$MeanErr$ and the correlation r were chosen as the outputs of the model for the sensitivity analysis. $MaxErr$ was also considered in order to have an information on the worst predictions, but it was not used as a criterion in the sensitivity analysis. $RMSErr$ was also calculated but not used as a criterion in the sensitivity analysis either. The computation of these performance statistics are detailed at the beginning of Sec. II.2.

The 625 combinations of parameter values were tested for the 5 experiments described above. The sensitivity indices were estimated using either r or $MeanErr$ as model output. The interactions between parameters and the relative influence of each parameter were studied using these indices. Afterwards, for each experiment, the independent parameters were varied independently to define the value(s) leading to the best predictions, whereas for the interacting parameters, these values were defined while varying the parameters simultaneously. The best predictions values were then compared across experiments in order to find a single common value for each parameter leading to good predictions across all experiments. A qualitative analysis of the predictions was also considered, to eventually help define the final parameter values if there were more than a single value leading to best predictions across experiments. Values leading to predictions conceptually wrong (e.g., missing a basic effect observed in the data) were excluded prior to this analysis. For each type of parameter, independent or interacting, if its original value (Collin and Lavandier, 2013; Cubick et al., 2018) was among the values leading to the best predictions, then this value was selected for the parameter because there was no relevant argument for a change. The definition of the best parameter values was done while keeping in mind which parameters were the most influential.

III.4.2 Results

The conclusions of the sensitivity analysis were similar when considering r or $MeanErr$ as model output. The sensitivity index values were different but the observed trends were same. Only the results obtained with $MeanErr$ are presented here. All the first order sensitivity indices and the second order sensitivity index between Ceiling and BE are displayed in Table III.b.

The most directly influential parameter (displayed in bold for each experiment in Table III.b) was Ceiling for CM1, CL1, CL4 and VL while it was BE for CM2. For example for CL4, the corresponding index was equal to 43%, meaning that 43% of $MeanErr$ variance (over the 625 predictions) was due to the variations of Ceiling. To say it differently, if the sensitivity analysis had been ran with a constant Ceiling and only the three others parameters were varied, then the $MeanErr$ variance would have been at least 43% lower. The only non-negligible second order sensitivity index was for the interaction between Ceiling and BE (17% on average across experiments, while the second highest second order index was limited to 1% on average).

The sum of all first order sensitivity indices and the second order sensitivity index between Ceiling and BE (sum of the two last lines of Table III.b), per experiment, led to rates higher or equal to 92% (including 100 % for CL1 and VL). In other words, across all experiments, the variation of $MeanErr$ was almost entirely due to direct impacts of the parameters and the interaction between BE and Ceiling. The few percent of variance left were split into the ten other sensitivity indices. From this observation, the choice of the final values to be used was done individually for SpecSamp and BU, but BE and Ceiling were considered together.

None of the experiments were discriminating to choose the SpecSamp value, in agreement with the fact that its first order sensitivity index was equal to 0% for four experiments. This

Exp.	CM1	CM2	CL1	CL4	VL
	1^{st} order indices (%)				
Ceiling	**58**	21	**48**	**43**	**68**
BE	20	**54**	28	8	12
BU	1	3	0	0	18
SpecSamp	0	0	0	15	0
Sum of 1^{st} order indices	79	78	76	66	98
	2^{nd} order indices (%)				
Ceiling/BE	19	14	24	26	2

Table III.b: First order indices, their sum and the second order index between Ceiling and BE for all experiments. The highest first order index for a given experiment is displayed in bold. Only the main interaction (between Ceiling and BE) is displayed here.

parameter had some limited influence only for the predictions of CL1, but in practice all values led to accurate predictions. Therefore, the original value used by Collin and Lavandier (2013) was kept unchanged (2 Gammatone filters per ERB, see Sec. III.4.4 concerning this choice).

Concerning BU, the predictions for CL4 and CL1 were not affected by changing BU values. For CM1 and VL, the longer the window duration the lower *MeanErr* ; and for CM2 the lowest *MeanErr* was reached with the longest window duration. Hence, the analysis suggests to change the value of BU from 24 to 400 ms (from 12 to 200 ms ERD). It was decided to set BU to 300 ms (150 ms ERD), a value not tested above but which corresponds to the midst of the binaural temporal windows reported in the literature (Culling and Summerfield, 1998; Culling and Mansell, 2013; Grantham and Wightman, 1979; Hauth and Brand, 2018), which ranged from 80 ms to 500 ms (from 40 to 250 ms ERD, see Introduction). The difference in predicted SRT using a BU of 300 rather than 400 ms was below 0.1 dB in each of the five experiments considered above.

The values for Ceiling and BE giving the best predictions were deduced by removing the values leading to inconsistent predictions. The tested values for Ceiling were: 8, 12, 16, 20, 24 dB ; for BE, they were: 16, 24, 80, 200, 400 ms (or 8, 12, 40, 100, 200 ms ERD). In CM1 and CM2, when BE was set to 80, 200 or 400 ms the model predicted SRTs with obvious deviation from the data for all values of Ceiling (e.g. no difference in predicted SRT for stationary and modulated noises in CM1). These prediction errors are considerably reduced with the shortest window durations, so that only the values 16 and 24 ms (8 and 12 ms ERD) remained for BE.

Model predictions for CM1 and CM2 led to conflicting results concerning the choice of Ceiling. The best predictions for CM1 were obtained for values equal to 20 or 24 dB, whereas the best predictions for CM2 were obtained for a Ceiling of 8 dB. This value was not considered further, because CM2 is the only experiment well predicted with a Ceiling of 8 dB. A Ceiling of 12 dB led a 3.2-dB overestimation of the SRTs in the conditions with one modulated noise in CM1. In this case, the model also predicted identical SRTs for the steady-state and modulated noise in the separated condition. The best model performances for CL1 were reached with a 12-dB Ceiling. Because the value of 12 dB also led to conflicting results between CL1 and CM1, it was not considered further. The remaining possible values for Ceiling after this first analysis were 16, 20 and 24 dB.

The original values of Ceiling and BE used by Cubick et al. (2018) have not been discarded, meaning that they did not lead to inconsistent model predictions. Ceiling and BE were thus set to these values, 20 dB and 24 ms, respectively.

III.4.3 Predictions of the revised model

The SRTs predicted with the revised model are presented as solid lines for each experiment in Figs III-2 to III-6. The predictions of the original model are plotted for comparison with

Exp.	r Orig.; Rev.	$MeanErr$ Orig.; Rev.	$RMSErr$ Orig.; Rev.	$MaxErr$ Orig.; Rev.
CM1	0.93; 0.96	1.3; 1.0	1.6; 1.3	3.5; 2.3
CM2	0.92; 0.94	1.0; 1.0	1.2; 1.0	2.4; 1.8
CL1	0.85; 0.85	0.5; 0.5	0.6; 0.6	1.3; 1.3
CL4	0.92; 0.93	0.6; 0.6	0.8; 0.8	1.6; 1.4
VL	0.87; 0.90	1.0; 0.9	1.2; 1.0	2.0; 1.8
Ewert et al. (2017)	NA; 0.91	NA; 1.4	NA; 2.0	NA; 7.1

TABLE III.C: Performance statistics of the original (Orig.) and revised (Rev.) model. $MeanErr$, $RMSErr$ and $MaxErr$ are computed in dB. The experiment of Ewert et al. (2017) was only used to validate the revised model.

dashed lines. On each figure, the performance statistics of the revised model are indicated (r, $MeanErr$, $RMSErr$ and $MaxErr$). A comparison of the performance statistics between the original and revised model is displayed in Table III.c, which shows that they are similar across experiments and both models predict accurately the data. $MeanErr$ and $RMSErr$ provide also comparable values for each experiment and model.

In CM1, the steady-state noise conditions (Fig. III-2) are well predicted with errors below 1 dB. The model overestimates the SRT in the presence of a single co-located modulated masker by 2.3 dB and it underestimates by 2.3 dB the SRT for the symmetrical condition involving 2 separated modulated noises. The revised model improved this last prediction by 1.2 dB due to the longer duration of BU.

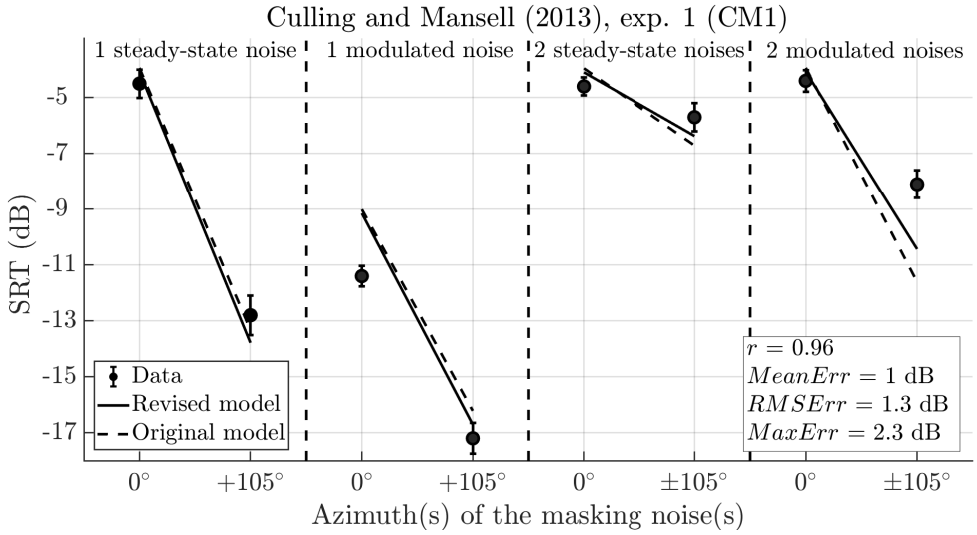


FIGURE III-2: Mean SRTs with standard errors across listeners measured in CM1, involving 1 or 2 noises, steady-state or modulated by a 10-Hz square wave (50% duty cycle, modulated out-of-phase if they were two maskers), simulated as originating from different azimuths (0° and +105° or ±105° if there were two maskers) in an anechoic environment. The target was always at 0°. Model predictions are displayed as a solid line for the revised model and as a dashed line for the original model. Model performance statistics are displayed only for the revised model.

In CM2, changing BU from 24 ms to 300 ms enables to better predict the influence of the modulation rate (between 1 and 5 Hz only) for the ITD-only conditions, and as a result also for the ILD+ITD conditions. Concerning the ILD-only conditions, it is important to note that the original and revised models predict exactly the same binaural ratios, i.e. the predictions of these conditions are not affected by the revision. Because the average prediction (which is different for the two models because of the other conditions) is scaled to the average SRT in the experiment, the resulting predicted SRTs are different. The model predicts SRTs increasing by about 1.5 dB above the 5 Hz modulation rate for the ILD-only conditions, while the data show a 0.6-dB difference.

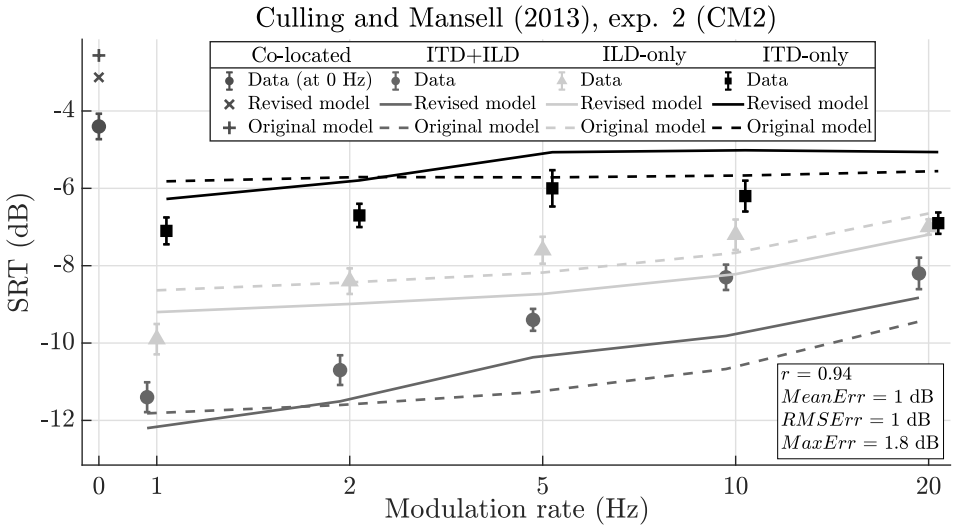


Figure III-3: Mean SRTs with standard errors across listeners measured in CM2. The target was always presented at 0° in the presence of two noises placed on both sides of the listener (±105°). The noises were modulated out-of-phase by a square wave at 5 modulation rates (1, 2, 5, 10, 20 Hz). Three types of HRIR were involved (ILD+ITD, ILD-only, ITD-only). One reference condition involved a steady-state noise co-located with the target (modulation rate of 0 Hz). Model predictions for the separated configuration are displayed as a solid line for the revised model and as a dashed line for the original model. The predictions related to the co-located configuration are plotted using a cross and a plus sign for the revised and original model, respectively. Model performance statistics are displayed only for the revised model.

For CL1 (Fig. III-4), there is no difference between the original and revised model, not surprisingly since target and masker were simulated in front of the listener, so that the influence of binaural unmasking was limited. In CL4 (Fig. III-5), the model predicts accurately all the conditions involving a single masker (i.e. black symbols), only the one with a co-located steady-state noise leads to an error of about 1 dB. For the conditions with two maskers (grey symbols), the model seems to predict about 1 dB more SRM than measured in the data.

The predictions for VL are quantitatively correct (Fig. III-6), suggesting that the model is able to predict the general trends measured in the data. The SRTs for the steady-state noise are better predicted than those for the 1-voice modulated noise. The model predicts a binaural unmasking advantage for the steady-state noise (difference between the black lines and the grey lines in the top panel) that was not observed in the data. In the bottom panel, the relative difference predicted between the no ITD/no tail conditions and the ILD+ITD conditions (grey solid lines and black solid lines, respectively) for a given spatial configuration does not correspond to the relative differences measured in the data. This means that the model is not able to completely predict the conflicting effects of having no ITD, i.e. no binaural unmasking,

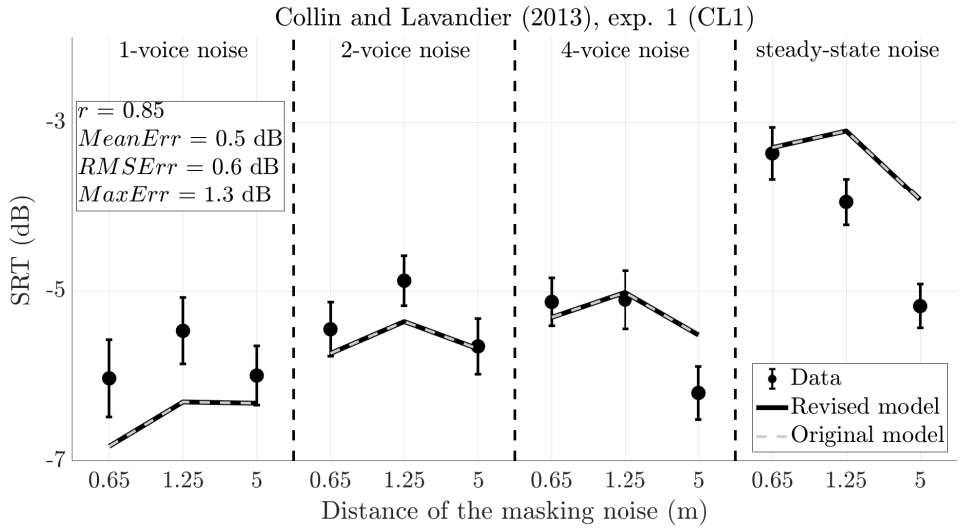


FIGURE III-4: Mean SRTs with standard errors across listeners measured in CL1. The target was at 0.65 m in front of the listener in a lecture hall. The noise was placed at three distances (0.65, 1.25, 5 m), also in front of the listener. Four types of modulation were used for the noise (steady-state, 1-, 2- or 4-voice modulated). Model predictions are displayed as a solid line for the revised model and as a dashed line for the original model. Model performance statistics are displayed only for the revised model.

and no reverberation tail, i.e. no filling in of the masker gaps.

III.4.4 Discussion

Considering the sensitivity analysis, it should be emphasized that the range of values over which the model parameters were varied will have influenced the magnitude of the sensitivity indices. For instance, a smaller range of Ceiling values could have led to a decrease of this parameter predominance. Inversely, a larger range of BU values could have led to higher sensitivity indices. The tested values were chosen based on previous results from the literature; but they should be kept in mind when considering the conclusions of the sensitivity analysis.

The only strong interaction between model parameters was observed for Ceiling and BE, the two parameters involved in the computation of the SNR at the better ear. The window duration BE sets the time constant for the model to analyze an amplitude modulation in a noise envelope and Ceiling sets the maximum value of the by-band SNR/masker modulation depth from which the band contribution to intelligibility is assumed to plateau. Conceptually, if the window duration is too long, the fast modulations will not be detectable. As a result, Ceiling will not be used in the calculation of the SNR at the better ear for those modulations. Conversely, if the window duration is sufficiently short for detecting the modulation, then Ceiling will be used in the calculation and will influence the model output. So it is not surprising that these two parameters interact.

The window duration BU used to compute the binaural unmasking advantage presented lower first order sensitivity indices than the window duration BE, probably for two reasons. First, it should be noted that the better-ear listening component of the model is influenced both by the ILD/better-ear effects, but also by the effects associated with masker modulations (dip listening). Across experiments, less conditions were tested in which binaural unmasking played a role compared to those in which better-ear/dip listening played a role (e.g. in the ITD-only conditions, the better-ear component of the model was still influenced by the differences in masker modulations). As a result, the model predicts more differences across conditions that are associated with the better-ear/dip listening component. Hence, it seems normal that the model is more sensitive to the parameter associated with this latter component. The second reason is

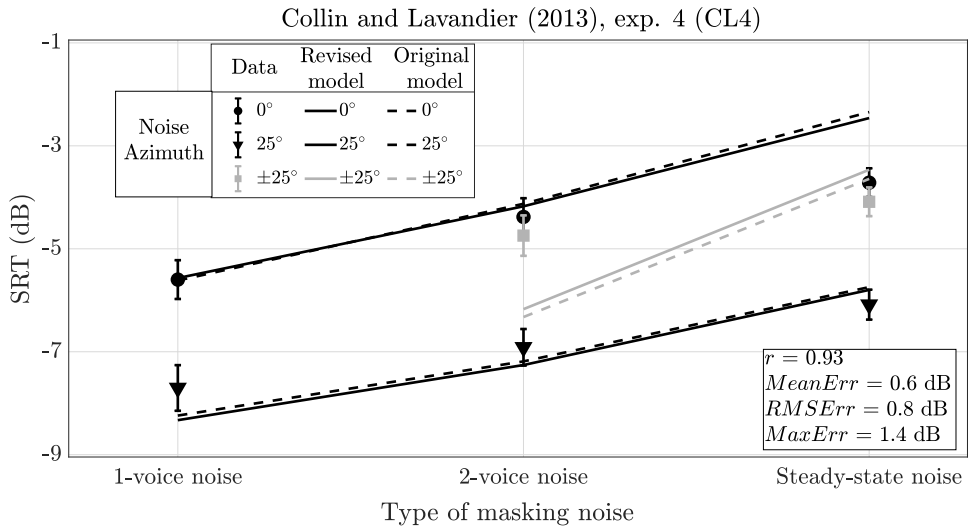


Figure III-5: Mean SRTs with standard errors across listeners measured in CL4. The target was at 0.65 m in front of the listener in a meeting room. The single masker was always at 0.65 m but tested at two azimuths (0° and 25°). Three types of noise were involved (1-voice modulated, 2-voice modulated or steady-state). Two noises (steady-state or two 1-voice modulated) were tested in two configurations (0° or ±25°, 0.65 m). The revised and original model predictions are plotted as a solid and a dashed line, respectively. Model performance statistics are displayed only for the revised model.

the following, as described in the previous section, the predictions were extremely far from the data when BE was set to the longer durations, only for the shorter BE the predictions described well the data. As a result, a considerable range of $MeanErr$ variations depended on BE. On the other hand, whatever the window duration BU was, the predictions were sufficiently close to the data, so that the range of $MeanErr$ variations induced by BU variations was narrower than the range induced by the BE variations.

Considering the final choice of the values used for the model parameters, first, binaural sluggishness is better taken into account by the model with a window duration BU equal to 150 ms (ERD). This allows to better predict the effect observed in the ITD-only condition in CM2 even if the amplitude of the effect is rather low (about 1-1.5 dB). The accuracy of the data can be criticized and this is why the literature was also used to justify the change of this parameter value. As discussed in Culling and Mansell (2013), binaural sluggishness would limit the auditory system ability to follow changes in binaural cues above 5 Hz. The duration of the window was increased in order to see this limitation in the model predictions, namely, the predicted SRTs increase up to 5 Hz and then asymptote. Furthermore, as mentioned in Sec. III.1, several previous studies measured the binaural temporal window using different methods (Culling and Summerfield, 1998; Culling and Mansell, 2013; Grantham and Wightman, 1979; Hauth and Brand, 2018). The obtained values were between 80 to 500 ms (40 to 250 ms ERD).

Hauth and Brand (2018) investigated the effect of binaural sluggishness using a different short-time binaural speech intelligibility model (Beutelmann et al., 2010, or see Sec. II.2.3.1). They designed an experiment in order to test the effect of binaural sluggishness on speech intelligibility. Stimuli were a steady-state noise for which IPDs were modulated sinusoidally between $-\pi/2$ and $+\pi/2$ at different rates between 0 and 64 Hz. Increasing the modulation rate led to higher SRTs for rates up to 4 Hz, above which the rate had no significant influence on the SRT. These results are consistent with the results of CM2. When modeling their own experiment, Hauth and Brand modified the EC processing of their model to introduce binaural sluggishness that influences the definition of the EC parameter. However, the EC stage *per se* is still applied on short-time signals (for detail of implementation, see Hauth and Brand, 2018). In the current model, the binaural unmasking advantage is estimated using signals whose duration is influenced by binaural sluggishness, resulting in longer signals than in the "revised" stBSIM.

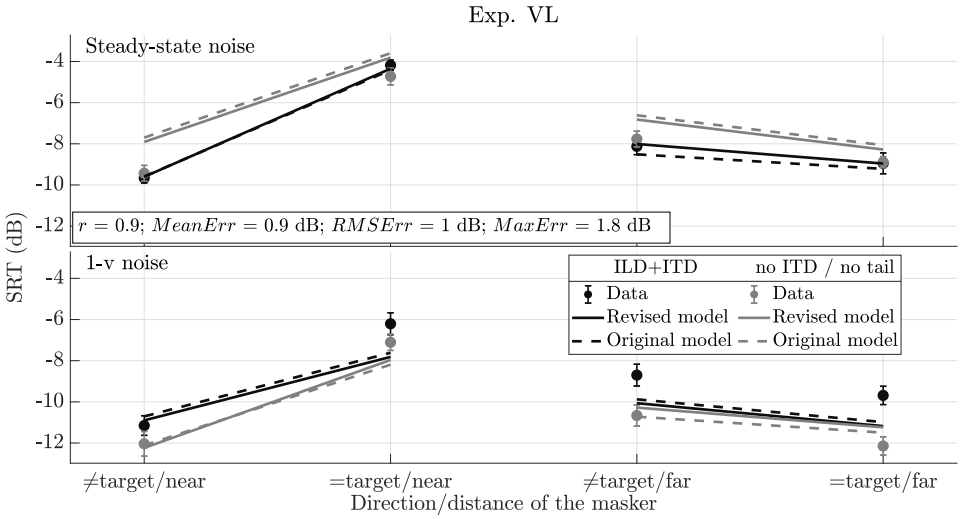


FIGURE III-6: Mean SRTs with standard errors across listeners measured in the present study (VL). The target was placed at 0.65 m, +25° from the listener (=target/near). The noise was steady-state (top panel) or 1-voice modulated (bottom panel). It was tested at two distances (near at 0.65 m and far at 5 m) and two azimuths (+25°/=target, -25°/ =target) in a room. Two types of BRIR were involved (natural BRIRs with ITD+ILD, SEIRs with no ITD/no tail). Solid lines present the revised model predictions, while dashed lines present the original model predictions. Model performance statistics are displayed only for the revised model.

Despite this discrepancy of binaural sluggishness implementation in the models, the duration of the binaural/EC window proposed by Hauth and Brand allowing to predict accurately their data was 200 ms (ERD), which is similar to the duration (BU) highlighted in the current study.

SpecSamp did not influence the model predictions in any of the 5 experiments tested at this stage. So if one is interested in saving computing time, it seems appropriate to reduce the spectral sampling of the model. Reducing it to as low as 2 filters per 5 ERB did not impair the predictions in the 5 experiments tested here. We choose to keep 2 filters per ERB for now, because computing time is not an issue for the current study and a better spectral sampling might be needed in future developments of the model (e.g. while considering harmonic maskers or hearing-impaired listeners).

Some studies have shown that speech intelligibility models could still lead to relevant predictions despite a loss of frequency resolution, which has to be understood as the accuracy to analyze the signals in the frequency domain. To change the frequency resolution in a model, the number and the bandwidth of the filters that analyze the signals have to be varied but keeping the same overall frequency bandwidth analysis. Kryter (1962), when developing the Articulation Index (AI, monaural speech intelligibility model), showed that its predictions using a 20-band, one-third-octave-band or octave-band method were in reasonable agreement. Steeneken and Houtgast (1980) developed and validated the STI (see Sec. II.2.1.1), with its computation done using an octave band method.

The conclusion of the present study is different because the results showed that the model predicted similar SRTs even if some frequency channels were not used for the computation while keeping the same filters (creating "holes" in the bandwidth in which the signals are analyzed). However, those conclusions lead to a common observation, which shows that a loss of spectral information in the signals, either by smoothing it (reduction of the frequency resolution) or not analyzing some frequency channels (reduction of the spectral sampling), still results in similar model predictions.

Regarding the Ceiling value, it has been set to 20 dB that is higher than the values implemented in the SII (ANSI S3.5, 1997) or in the AI (Kryter, 1962), +15 and +18 dB SNR, respectively. It means that the current model considers that the full target intelligibility is

reached at a higher SNR. Studebaker and Sherbecoe (2002) showed that increasing the SNR up to 29 dB could still improve target intelligibility. Such a high value of Ceiling does not seem appropriate in the proposed model. Collin and Lavandier (2013) introduced Ceiling to the model. They found that a value of 10 or 15 dB reduced prediction errors. Cubick et al. (2018) needed a Ceiling of 20 dB to optimize predictions. Implementation differences between both models (see Sec. III.3) may account for the different Ceiling values.

The values of the window duration BE used to compute the better-ear component of the model that predicted well most conditions tested here were 16 and 24 ms (8 and 12 ms ERD). These values are within the range 8 to 13 ms (ERD) of the measured monaural temporal resolution (Moore et al., 1988; Plack and Moore, 1990). The tested binaural window durations (i.e. 80, 200, 400 ms or 40, 100, 200 ms ERD) provided inconsistent predictions in some conditions. The model was not able to predict the advantage of listening in the masker dips when the temporal resolution of the better-ear component was not sufficient (i.e. when the window duration was too long). For instance, in CM1 a BE of 200 ms provided the same predicted SRT for the steady-state and modulated noises. Taking a too-long temporal window triggers an amplitude modulation smoothing in the model, so that the modulated masker appeared as a steady-state masker. Consequently, the window duration BE has to match a monaural time constant. It should be noted that Collin and Lavandier (2013) as well as Beutelmann et al. (2010) also used a temporal resolution of 24 ms. It corresponds to the best frequency-independent duration used in the monaural model of Rhebergen and Versfeld (2005).

The values retained for BE differed from Culling and Mansell's conclusion, which stated that better-ear listening is a mechanism affected by binaural sluggishness, because in CM2 there was an influence of the required ear-switching rate up to 5 Hz. Although a monaural time window is required for the proposed model, it does not mean that better-ear listening is a "double" monaural mechanism, which is not influenced by binaural sluggishness and across-ear switching. It just means that the current implementation of the model does not allow for predicting this effect of sluggishness on better-ear listening. Culling and Mansell (2013) concluded that better-ear listening is binaural because the listener has to choose which ear is more beneficial for listening to the target; but also that the monaural behavior of each ear allows for listening in the dips. So there may be two time constants to consider for modelling better-ear listening. Modelling the effect of across-ear switching on better-ear listening is not straightforward and not implemented here. The present study however shows that better-ear listening cannot be simply modeled using a binaural temporal window. The monaural temporal resolution is required to predict the benefit associated with fast masker modulations.

III.5 Validation of the revised model

III.5.1 Predictions

The model predictions in the 5 HRIR conditions of Ewert et al. (2017) are plotted in the panels of Fig. III-7. The SRTs were scaled using the mean SRT across all 40 conditions (i.e. the scaling was done only once for all panels rather than independently for each panel), in order to observe whether the model could predict the differences across HRIR conditions. The model performance across all conditions led to r equal to 0.91, $MeanErr$ of 1.4 dB, $RMSErr$ of 2.0 dB and $MaxErr$ of 7.1 dB. The correlation r and $MeanErr$ are similar to those obtained for the other experiments presented above. $MaxErr$ is considerably larger due to a single data point (last panel of Fig. III-7). The performance statistics were also computed separately for each HRIR condition and are displayed in the corresponding panels of Fig. III-7.

The general pattern of the predictions and the model performances are similar for the natural ITD+ILD, ILD-only and IMBM conditions (first, second and fourth panels, respectively). The solid black and grey lines represent the predictions for the co-located and separated conditions, respectively. The correlations r are above 0.92 and $MeanErr$ around 1 dB. $MaxErr$ is obtained for the separated conditions with the 1-voice Freq. Inc. noise. The differences in SRTs produced by the different types of masker modulation are well predicted, for both spatial configurations, except for the 1-voice Freq. Inc. modulation in all HRIR conditions and the sinusoidal modulation in the separated IMBM condition, where the differences between the observed and predicted SRTs are around 2-3 dB.

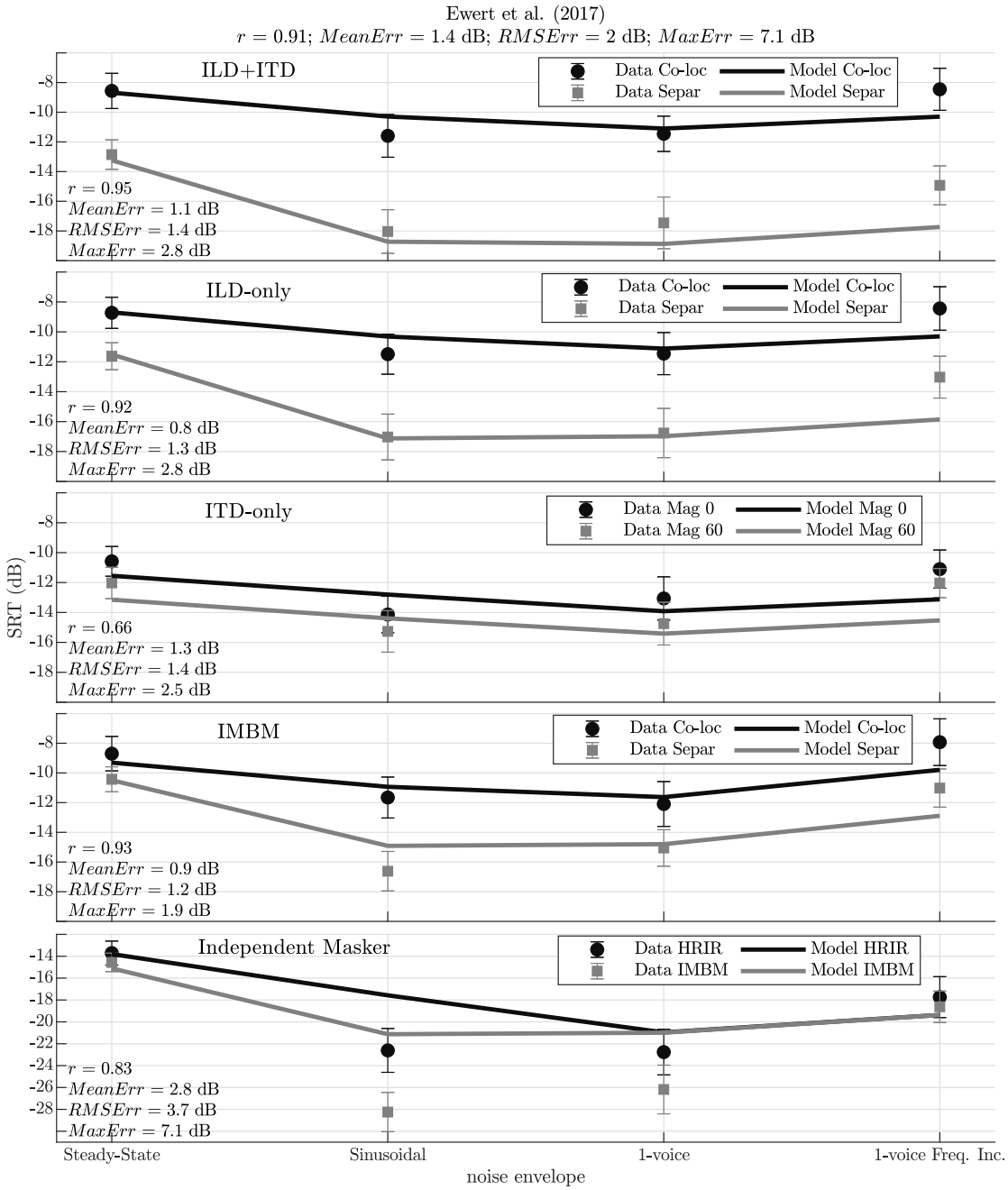


FIGURE III-7: Mean SRTs with standard deviations across listeners measured by Ewert et al. (2017). SRTs are plotted as a function of the noise modulation type (steady-state, sinusoidal, 1-voice, 1-voice frequency incoherent). Each panel corresponds to a given type of HRIRs (ITD+ILD, ILD-only, ITD-only, IMBM, Independent Maskers). For the first four panels, two spatial configurations were tested : while the target was at 0°, the two noises were placed in front (at 0°) or on each side (±60°) of the listener (plotted in black circles and grey squares, respectively). For the last panel, two different HRIRs (natural and IMBM) at 0° were used to create the independent maskers represented with black circles and grey squares, respectively. The model predictions are plotted in solid black and grey lines in all panels. The model performance statistics across all conditions are indicated in the title, and the performances for each HRIR type are displayed in the corresponding panel.

The ITD-only predictions related to the Magnitude 0/60 are plotted on the third panel, with solid black and grey lines, respectively. The trends across masker modulations are not well predicted (r = 0.66) although $MeanErr$ is close to 1 dB. The 2.5-dB $MaxErr$ occurs again for the separated configuration with the 1-voice Freq. Inc. noise. The difference between the black and grey lines is almost equal to the difference between the solid black and grey symbols, indicating that the model predicts the influence of spectral coloration.

The predictions for the independent masker conditions are shown in the last panel. The steady-state noise conditions are well predicted, while the SRTs for the modulated noises are less well predicted. The SRTs are overestimated for the sinusoidal and 1-voice modulation, while they are underestimated for the 1-voice Freq. Inc. modulation. The model does not predict any difference between the HRIR and IMBM for the 1-voice and the 1-voice Freq. Inc. modulation (while there is a difference in the data). $MaxErr$ occurs for the independent IMBM condition with sinusoidal modulation. $MeanErr$ is equal to 2.8 dB for this HRIR condition, which is the worstly predicted.

III.5.2 Discussion

First, the overall model performance on this experiment is relatively good, and comparable to the stBSIM performance (Ewert et al., 2017). The performance statistics of the two models cannot be compared because the prediction errors of the stBSIM were largely increased due to mispredictions of informational masking. Those conditions were not even attempted for here, because they cannot be described with our only-energetic-masking model. The proposed model is able to predict the difference across HRIR conditions, particularly the differences between natural ITD+ILD, ILD-only and ITD-only conditions. In other words, each model component is able to predict the effect of its associated binaural cue (ITD or ILD) and the combination of the two components leads to accurate predictions of the natural (ILD+ITD) HRIR conditions.

Predictions for the 1-voice Freq. Inc. noises always underestimate the measured SRTs. The model predicts too much advantage when the listeners were listening in gaps that were incoherent across frequency bands. Compared to the conditions with 1-voice modulated noises, the predicted SRTs are always higher, indicating that the model predicts a detrimental effect of the incoherence across frequency, but not enough to match the data. This suggests that the auditory system is not able to use incoherent amplitude modulation as well as coherent amplitude modulation, which would mean that the unmasking auditory mechanisms (i.e., better-ear listening and binaural unmaking) are not independent from each other across frequency as assumed in the model. This discrepancy was also observed in the stBSIM predictions (Ewert et al., 2017), which would confirm this explanation because it also applied independent unmasking mechanisms across frequency bands.

Model predictions are less accurate for differences amongst the ITD-only conditions, r equals 0.66 indicates that the model is not able to predict correctly the trends in the data across masker types. The model predictions show a pattern similar to the stBSIM predictions. The SRTs for the steady-state, 1-voice and 1-voice Freq. Inc. noises are underestimated (by 1 to 2 dB), while the SRTs for the sinusoidal masker modulation are overestimated (by 1 dB).

The model predicts correctly only half of the data measured with the independent HRIR/IMBM conditions. The predictions for the steady-state and 1-voice Freq. Inc. modulated noises match the data but the predicted SRTs for the sinusoidally and 1-voice modulated noises are largely overestimated, leading to an error of 7.1 dB. The stBSIM was more accurate to predict the magnitude of the variations across masker types. $MaxErr$ was probably below 5 dB, occurring for the SRT measured with 1-voice Freq. Inc. modulated noises and the independent HRIR. Conversely, the stBSIM predicted higher SRTs with modulated noises for the independent IMBM conditions than for the independent HRIR conditions. Therefore, the stBSIM and the present model show limits (different for each model) to predict the influence of these types of artificial HRIRs.

III.6 General discussion

In the end, the present study led to a single change of the parameter values used in the original model. The original values were inspired from the literature when the model was developed (Collin and Lavandier, 2013). Only Ceiling was roughly tested and then fixed by Cubick et al.

(2018). The present study confirms that these values are indeed required for optimal predictions. The value of BU has been modified to take into account the effect of binaural sluggishness allowing to better predict the influence of the modulation rate on binaural unmasking in CM2. The influence of this revision is of course limited here because it models an effect that is not dominating in the experiments considered.

Despite the model's revision, some conditions are still not well predicted, as is the case for the effect of reverberation when it fills the masker's gaps. For instance in Fig. III-6, the difference between model predictions (black lines) on each panel, which represents the dip-listening advantage, is around 1.5 dB higher than the difference in the data (black circles). Therefore, the model overestimates the dip-listening advantage even if the trends are well predicted. For the conditions without reverberation tail (grey squares and grey solid lines), the dip-listening advantage is better predicted. Hence, the current model does not fully take into account the negative effect of reverberation filling in the gaps in the masking noise. Earlier, Beutelmann et al. (2010) observed a similar behavior for their model.

The model overestimates the SRT for the co-located modulated noise in CM1 (Fig. III-2) while the corresponding separated condition is well predicted. This might be explained by the predictability of the dip occurrences within the masker. Fogerty et al. (2018) turned on and off a noise at different rates, with a 50% duty cycle, roughly resulting in a masker modulated by a square wave at different rates. They showed that listeners were able to benefit from the predictability of the dip occurrences for gating rates below 16 Hz. Collin and Lavandier (2013) found similar results concerning the predictability of masker dips using noise modulated by a broadband speech envelope. Culling and Mansell's masker was modulated by a 10-Hz square wave in CM1, so listeners were probably able to at least partly benefit from the predictability of the masker dips. The difference observed between data and prediction could be due to this effect, which is not taken into account by the model, the parameters of which were set to predict unpredictable speech modulations (in CL1, CL4 and VL).

For the symmetrical configurations with modulated maskers on both sides of the target (e.g. in CM1, CM2 (ILD+ITD), CL4), the model predicts more better-ear glimpsing and/or binaural unmasking than measured in the experiments. Increasing the duration of the temporal window BE used for computing the SNR at the better-ear in order to simulate binaural sluggishness and the across-ear switching cost did not produce better predictions in the present study. A future update of the better-ear listening model component — which could take into account binaural sluggishness along with the monaural ability for listening in the dips — could improve the predictions for these conditions.

Some other effects could have been tested in the present study and potentially added in the model, which might further improve its prediction accuracy, even if being detrimental to its simplicity. While adapting the monaural SII for fluctuating noise using temporal windows to decompose the signals, and inspiring the binaural models presented here and by Beutelmann et al. (2010), Rhebergen and Versfeld (2005) showed that their best predictions were obtained with frequency-dependent durations for the temporal window. The 12 ms-ERD used in the present model comes from their best value for a frequency-independent window, but it is only an approximation of a more complex frequency-dependent decomposition of the signals. Also, Rhebergen et al. (2006) later implemented forward masking in their model. This additional component led to better predictions in the case of a periodically modulated noise. The shape of the temporal windows used in the present model (Hann windows) could have been varied. Culling and Summerfield (1998), Moore et al. (1988), as well as Plack and Moore (1990) indicated that the shape of this window is probably asymmetric and depends on the frequency and level of the stimulus. Finally, Beutelmann et al. (2009) and Kolarik and Culling (2010) demonstrated that binaural auditory filters are probably wider than monaural auditory filters. This feature could be incorporated and tested in the proposed model, particularly for the prediction of the binaural unmasking advantage.

The revised model proposed here provides predictions similar to the original models it is based on (Lavandier and Culling, 2010; Jelfs et al., 2011; Lavandier et al., 2012; Collin and Lavandier, 2013) and other models proposed in the literature (Beutelmann et al., 2010; Wan et al., 2014), with r ranging from 0.85 to 0.96 (across experiments) and $MeanErr$ between 0.5 and 1.4 dB. The value of only one parameter was changed compared to the model of Collin and Lavandier (2013). This change allows to take at least partly into account the effect of binaural sluggishness on binaural unmasking. More importantly, all model parameters have been thoroughly tested,

and it was demonstrated that the parameter values proposed are those giving the best results. The model has been validated on three speech corpora (German, English and French), in anechoic and reverberant rooms, in the presence of different number of maskers and different types of masker modulations (steady-state, speech modulated or periodically modulated), with maskers placed at various azimuths and distances from the listener. In total, 60 conditions (CM1, CM2, CL1, CL4, VL) were used to set the value of the four model parameters, 20 of which (CL1, CL4) were previously used to validate the original version of the model. The revised model, using the new set of parameter values, was validated with an additional 40 conditions (Ewert et al., 2017). Thus, the robustness of the model has been improved through this study and it is more in line with the literature by the implementing a window accounting for the binaural sluggishness. The next chapter addresses the limitations of the model proposed by Lavandier et al. (2018) considering the outcomes of the current study.

The proposed model is available to the community. A code can be downloaded here: https://mathieulavandier.wordpress.com/home/models/.

IV

A binaural model implementing an internal noise to predict the effect of hearing impairment on speech intelligibility in non-stationary noises

This study proposed a revision of the model proposed by Lavandier et al. (2018), which provides a single binaural model version that is able to predict speech intelligibility in noise for both NH and HI listeners. To do so, an internal noise was implemented in the model to simulate the listener's hearing loss. The internal noise was modelled with two components: one related to hearing threshold elevation and the other related to the effect of external stimulus level on speech intelligibility. The revised model was tested on three datasets from the literature. The content of this chapter is published in the Journal of the Acoustical Society of America (Vicente et al., 2020).

IV.1 Introduction

The binaural model developed by Lavandier et al. (2018) was designed to predict speech intelligibility in noise for NH and HI listeners. It accurately predicted NH and HI data measured in two (anechoic) spatial configurations, but applied a model parameter (associated with the broadband level of the internal noise modelling the listener's hearing loss) that had to be chosen differently for the NH and HI listeners to obtain accurate predictions; effectively resulting in two different model versions. Furthermore, their approach divided the listeners only in two groups (NH and HI), even though there are many different degrees (and types) of hearing loss (e.g., see Fig. II-7), which would again require separate model versions.

In the present study, the model developed by Lavandier et al. (2018) is revised, so that a single model can be used to predict speech intelligibility for listeners with various degrees of hearing loss instead of separate models. It is accomplished by modifying the implementation of the internal noise, which consists now of two components: one related to elevated hearing thresholds and the other related to the effect of external stimulus level on intelligibility. This approach is in line with the findings from Bernstein and Trahiotis (2008), who provide an overview of the literature on the concept of internal noise before conducting experiments to measure detection thresholds for a tone in noise to characterize the internal noise in NH listeners. In line with their literature review, their results suggest that the internal noise would consist of two components. The first component is stimulus-independent and determines the absolute threshold (i.e., serving as "noise floor"). The second component is stimulus-dependent, with its level increasing when the external noise level increases following a dB-for-dB rule.[5]

In order to measure the internal noise floor, experiments of detection of a tone in quiet and detection of a tone in noise can be used (Bernstein and Trahiotis, 2008; McFadden, 1968). These experiments consist of measuring the level at which a target tone is detected using an adaptive procedure. The first step is to compare the masking level difference in quiet between a monaural (or diotic/in-phase) tone and an out-of-phase tone. Then, the data from Robinson and Jeffress (1963) — measuring masking level difference as a function a noise interaural correlation — are used to derive the correlation of the internal noise floor. Afterwards, the masking level

[5]It should be noted that they drew these conclusions using a fixed external noise level (within a condition) and varying the level of the tone such that the external noise level determined the overall stimulus level.

differences between a homophasic and antiphasic tone in noise[6] at different low sensation levels can be used to find the level of the internal noise floor. However, this requires to assume that the combined (internal+external) noise coherence, so-called effective coherence, is equal to the mean of the two noise interaural coherences weighted by their power level. So that, one can find the level of the internal noise floor that leads to the measured masking level difference using the curves of Robinson and Jeffress (1963) and the effective coherence. The stimulus-dependent internal noise can be characterized by measuring detection of tone in noise. This part of the internal noise limits the efficiency of binaural processing at high sensation level, which means that when the level of the masker is increased by, say X dB, the level of the tone to be detected is also increased by X dB (because the internal noise follows a dB-for-dB rule).

Beutelmann models (Beutelmann and Brand, 2006; Beutelmann et al., 2010) predicting binaural speech intelligibility in noise consider an internal noise that is spectrally shaped on the listener's audiograms. The internal noise is then added to the external noise, thus, modifying the external noise characteristics. Plomp (1978) proposed to consider the listener's hearing loss in two components, one reflects the fact that signals are attenuated and the second is related to the loss of coding of these signals. This behaviour is similar to the internal noise definition of Bernstein and Trahiotis (2008). With this approach, Plomp (1978) proposed a formula to derive SRTs as a function of external noise level and hearing loss.

The modification applied to the model of Lavandier et al. (2018) consists in implementing an internal noise floor with a stimulus-dependent internal noise and considering the listener's hearing loss in two kind of losses. The revised model is optimized and verified here using data from three experiments involving NH and HI listeners (Rana and Buchholz, 2016, 2018a,b), as well as data measured with only NH listeners (Collin and Lavandier, 2013; Lavandier et al., 2012) to verify its backward compatibility. Besides addressing the model performance, the discussion provides an in-depth analysis of the internal noise implementation and a comparison with other models proposed in the literature that consider an internal noise.

IV.2 Model description

IV.2.1 Original model developed for NH listeners

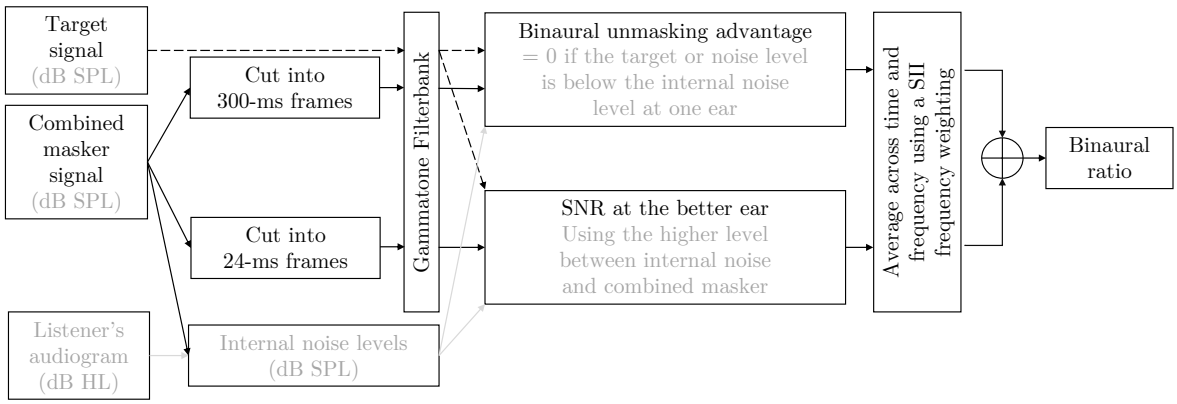


FIGURE IV-1: Block diagram of the proposed model. The original model designed only for NH listeners is presented in black (Vicente and Lavandier, 2020), the modifications associated with the extension to HI listeners are highlighted in grey.

A block diagram of the proposed speech intelligibility model is shown in Fig. IV-1, with the components of the original model (Collin and Lavandier, 2013; Vicente and Lavandier, 2020)

[6]Homophasic or antiphasic tone in noise mean here that the tone has the same phase as the noise or the tone is out-of-phase with the noise.

highlighted with the black font. The target and combined masker signals at the listener's ears, equalized to the same broadband level (mean across ears), are taken as inputs to the model. The target characteristics (i.e., magnitude spectra at each ear and ITDs) are averaged across time, computed only once on the whole signal to get their long-term values. This avoids that a short pause between words leads to a very low SNR and thus a poor predicted intelligibility, even though it provides relevant information for the listener. The masker characteristics are computed as a function of time for the model to predict the ability of a listener to understand speech in the dips of the masker's envelope ("dip listening", Festen and Plomp, 1990).

Based on the incoming signals, the SNR at the better ear and the binaural unmasking advantage are computed per time frame and frequency band combining the target's long-term characteristics with the masker's short-term characteristics (i.e., magnitude spectra at each ear, ITDs, and interaural coherence). To compute the SNR at the two ears, the masker signals are segmented using 24-ms half-overlapping Hann windows, so that "monaural" dip listening can be predicted. To compute the binaural unmasking advantage, 300-ms half-overlapping Hann windows are used so that binaural sluggishness can be taken into account (Vicente and Lavandier, 2020; Hauth and Brand, 2018). The frequency analysis is realized by applying two Gammatone filters per equivalent rectangular bandwidth, covering a frequency range from 30 Hz to the closest gammatone filter center frequency below the half of the sampling frequency, i.e., 9.6, 19.9 and 25.1 kHz for the signals sampled at 20 kHz (Lavandier et al., 2012), 44.1 kHz (Rana and Buchholz, 2016, 2018a,b) and 48 kHz (Collin and Lavandier, 2013), respectively.

The better-ear SNR is computed choosing the higher SNR between the left and right ears, with a ceiling SNR at 20 dB to avoid that the SNR goes to infinity in the masker's dips. The binaural unmasking advantage is computed by applying the binaural masking level difference formula from Culling et al. (2005). Their values are averaged across time, integrated across frequency using a SII weighting (derived from the band importance function in Table 1 of ANSI S3.5, 1997), and added to obtain a binaural ratio. Differences between binaural ratios can be directly compared to differences between SRTs measured in listening tests.

In order to derive predicted SRTs, the binaural ratios are first inverted, so that a higher binaural ratio reflects better speech intelligibility. Then, for each experiment, they are offset to fit the data by subtracting their mean and adding the average measured SRT across conditions and listeners[7], which was chosen as a reference. The predicted SRTs resulting from this transformation have the same mean value across conditions as the measured SRTs, for a given experiment.

IV.2.2 Extension of the model to HI listeners

In order to quantify the effect of reduced audibility, Lavandier et al. (2018) applied a number of changes to the original model of Collin and Lavandier (2013), using the implementation suggested by Cubick et al. (2018). These changes are highlighted with the grey font in Fig. IV-1. First, the model input signals are calibrated to the sound level (in dB SPL) used during the experiment (i.e. the sound level of the masker here, that was fixed during the adaptive measurements). This means that any amplification applied to the stimuli in order to compensate for the listener's hearing loss is considered by the model. To take into account the hearing loss, an internal noise is implemented at each ear with a spectrum that matches the individual audiograms. To compute the SNR at the better ear, the SNR at each ear is determined by the higher level between the external and internal noise and limited to 20 dB. The highest SNR across ears is selected as the better-ear SNR. The binaural unmasking advantage is computed only if the masker and target levels are above the internal noise levels at both ears.

The differences applied in the present study to the model of Lavandier et al. (2018) are solely related to the implementation of the internal noise. In this regard, three assumptions are made. (1) The overall level of the external stimuli is approximated by the known masker level, assuming that the broadband SNR is below 0 dB. (2) The hearing loss of the listener can be split into proportions η and $1-\eta$ to reflect the different contributions of outer hair cell (OHC) and inner hair cell (IHC) loss, with η being identical at all frequencies and for all listeners.

[7]When NH and HI listeners are involved, the SRTs are averaged across listeners within each group and then averaged across conditions and groups, giving the same weighting to each group regardless of their number of listeners. The same approach is used for the binaural ratios when removing their mean.

(3) The maximum value allowed for the estimated OHC loss is 57.6 dB (Moore and Glasberg, 2004) and the loss above this value contributes solely to the estimated IHC loss. The estimated OHC and IHC losses are interpolated at the model's center frequencies between the lower and upper center frequency of the audiograms (250 Hz and 8,000 Hz, respectively) and extrapolated otherwise, using a logarithmic frequency scale. The level (in dB SPL) of the internal noise N_{int} in each frequency band is then calculated using the following formula:

$$N_{int}(n,N) = \Gamma(n) + \boldsymbol{\eta} T(n) + 10log_{10}(10^{\frac{\boldsymbol{B}}{10}} + 10^{\frac{N-\boldsymbol{N_{lim}}+(1-\boldsymbol{\eta})T(n)}{10}}) \quad \text{(IV.1)}$$

where n is the center frequency of the n$^{\text{th}}$ frequency band and N is the long-term broadband level of the external masker averaged across ears.[8] Given that the model takes into account the listener-individual amplification applied in the experiment, the term N can vary across listeners (if they have different hearing loss profiles leading to individual amplification). The term $T(n)$ refers to the standard pure-tone audiogram in dB HL, thus, $\eta T(n)$ and $(1-\eta)T(n)$ are the estimated contributions in dB HL related to OHC and IHC loss, respectively. The function $\Gamma(n)$ is the transformation to convert the hearing loss from dB HL to dB SPL. This transformation results from the sum of the reference equivalent sound pressure levels for the THD 39 headphones used when measuring the audiograms (ISO 389-2, 1994) and nominal values for the transformation from 6 cc coupler to ear drum levels (Bentler and Pavlovic, 1989). The transformation follows the same interpolation/extrapolation as the audiograms (interpolated between 200 Hz and 6 300 Hz, the range where values are available) to derive its values at the center frequencies of the model. The frequency-independent free parameters B, N_{lim} and η (highlighted in bold in Eq. IV.1) were systematically varied in the present study for the experiments involving HI listeners, before being set at -10 dB, 83 dB and 0.7, respectively (see Sec. IV.2.3).

The proposed implementation of the internal noise models a behaviour that resembles the different effects of OHC and IHC loss on speech intelligibility. It is thereby assumed that the OHCs are mainly related to the audibility of the incoming sounds, while the IHCs are mainly related to their coding (e.g., Moore and Glasberg, 2004). Eq. IV.1 at low sound levels N (i.e., $N - N_{lim} + (1-\eta)T(n) \ll B$) can be simplified to:

$$\begin{aligned} N_{int}(n,N) &\sim \Gamma(n) + \eta T(n) + 10log_{10}(10^{\frac{B}{10}} + \cancel{10^{\frac{N-N_{lim}+(1-\eta)T(n)}{10}}}) \\ &\sim \Gamma(n) + \eta T(n) + B. \end{aligned} \quad \text{(IV.2)}$$

Hence, for soft sounds and poor audibility, only the part of the audiograms related to the OHCs ($\eta T(n)$) is considered. For high sound levels N (i.e., $N - N_{lim} + (1-\eta)T(n) \gg B$), Eq. IV.1 can be simplified to:

$$\begin{aligned} N_{int}(n,N) &\sim \Gamma(n) + \eta T(n) + 10log_{10}(\cancel{10^{\frac{B}{10}}} + 10^{\frac{N-N_{lim}+(1-\eta)T(n)}{10}}) \\ &\sim \Gamma(n) + \eta T(n) + (1-\eta)T(n) + N - N_{lim} \\ &\sim \Gamma(n) + T(n) + N - N_{lim} \end{aligned} \quad \text{(IV.3)}$$

This indicates that, at high sound levels, in addition to the audibility issue related to the OHC loss, intelligibility is further impaired by the IHC loss.

IV.2.3 Model evaluation

The proposed model was evaluated on three datasets involving NH and HI listeners (Rana and Buchholz, 2016, 2018a,b), which considered the effects of hearing loss, sensation level, spatial configuration of target and maskers, and masker temporal envelopes. Three indices of model performance were considered: the correlation r, $MeanErr$ and $MaxErr$ (see Sec. II.2 for computation details).

To quantify the improvement in prediction performance obtained with the current internal noise implementation compared to the implementation proposed by Lavandier et al. (2018),

[8] The long-term broadband level is preferred in order to provide an internal noise with the same spectral shape as the audiograms. This would not have been possible if the external noise level per band had been considered.

their models were applied here as reference ("Lav18 models"). The parameter value used to set the level of the internal noise for the NH listeners was -11 dB and -22 dB for the HI listeners (Lavandier et al., 2018). Compared to the original paper, instead of using a separate reference to convert the binaural ratios in predicted SRTs for each group of listeners, a common reference is used here for both groups[7], in order to evaluate whether the difference across groups can be predicted.

The backward compatibility with previous model versions was verified using two datasets involving only NH listeners. To compare predictions for a stationary noise masker, the model version and data from Lavandier et al. (2012) were considered ("Lav12 model"). To evaluate the effects associated with masker envelope modulations, the model version and data from Collin and Lavandier (2013) were considered ("Col13 model"). For a fair comparison, the components that are similar across models were implemented in the same way.

The Lav12 model follows the implementation presented in Sec. IV.2.1, except that no time frame analysis is applied, i.e., the better-ear SNR and the binaural unmasking advantage are computed on the long-term signal characteristics. The Col13 model follows the implementation presented in Sec. IV.2.1, but using the same 24-ms frame to compute the better-ear SNR and the binaural unmasking advantage (instead of using 24-ms and 300-ms frames proposed later by Vicente and Lavandier, 2020).

For all model predictions, the target signal was created by averaging between 60 and 128 sentence waveforms.[9] All sentences were truncated to the shortest sentence duration before averaging. The duration of each masker signal was at least 2 minutes. All signals from the experiment of Rana and Buchholz (2016, 2018a,b) were both convolved with the impulse response of the (equalized) headphones used for data collection and measured on a 4128C Bruel&Kjaer head and torso simulator. Target and masker signals were calibrated to the fixed sound level averaged across ears used for the masker in the experiments.[10]

In order to find the best combination of parameters in Eq. IV.1, the free parameters B, N_{lim} and η were varied within the ranges [-16;-8] dB, [65;85] dB and [0.6;0.9] (Pieper et al., 2018), with the aim of simultaneously minimizing $MeanErr$ and $MaxErr$ and maximizing r. This optimization of the model performance was done only using the three datasets involving both NH and HI listeners (Rana and Buchholz, 2016, 2018a,b), and the best predictions were obtained for: B = -10 dB, N_{lim} = 83 dB and η = 0.7.[11] The Spearman's rank correlation coefficient r_s was also computed between data and predictions for each experiment but not used as a criterion in the optimization stage.

IV.3 Results

The results shown below compare the speech intelligibility data taken from the literature with the corresponding model predictions (see Sec. IV.2.3). For the data, only brief overviews of the experimental designs are presented, the detailed descriptions and analyses are available in the original publications.

IV.3.1 Dataset involving NH and HI listeners

The proposed model was validated using three different datasets that shared some common methods. They were measured with native English speakers who had either normal hearing (hearing loss < 15 dB HL up to 6 kHz) or sensorineural hearing loss with less than 10-dB-HL difference across ears at any audiometric frequency up to 4 kHz (symmetric hearing loss). Moreover, the stimuli were anechoic and presented binaurally using equalized headphones. The target speech was from a BKB-like corpus (Bench et al., 1979), consisting of 80 lists of 16

[9] The exact numbers of sentences used to model the experiments of Collin and Lavandier (2013); Lavandier et al. (2012); Rana and Buchholz (2016, 2018a,b) are 60, 120 (for each type of target), 120, 128 and 128, respectively.

[10] When using the Lav12 and Col13 models, the input signals are calibrated to the same level, which does not need to be the actual level used in the experiment because the output of these models is independent of the absolute level of their inputs.

[11] The actual value of η can be lower than 0.7 because the estimated OHC loss is limited to 57.6 dB. Any listener who presents a loss above 82 dB HL (57.6/0.7) of hearing threshold would have an estimated proportion of OHC loss below 0.7.

meaningful sentences containing between 4 and 7 words. It was always presented from the front of the listener simultaneously with two SSNs or two noise-vocoded speech maskers (VSs, envelope modulated SSNs) either co-located with the target or spatially separated at ± 90°. The frontal position was simulated by convolving the anechoic stimuli with a head-related transfer function for frontal incidence averaged across ears, resulting in diotic listening. The noises were presented at different sensation levels and the relative target level was adapted to derive the SRTs.

The predictions for the NH listeners of Rana and Buchholz (2016) and Rana and Buchholz (2018b) were computed simulating a hearing loss at 0 dB HL at all frequencies for both ears because the audiograms were not available. This was not the case in Rana and Buchholz (2018a), where the individual audiograms for NH and HI subjects were available at each ear and used as model inputs.

The predictions of the Lav18 models are not shown here because the figures would have been overloaded. However, the predictions are available in Appendix B and the statistic performances and the main limitations (when applicable) are reported below for each experiment.

IV.3.1.1 Experiment 1 of Rana and Buchholz (2016)

Ten young NH listeners aged between 23 and 42 years (mean age of 31.1 years) and 10 older HI listeners aged between 49 and 77 years (mean age of 66.9 years) were involved in this experiment. The four-frequency (0.5, 1, 2, 4 kHz) average hearing loss (4FAHL) and ± 1 standard deviation of the HI listeners was 37.8 ± 7.1 dB HL. Both types of noise, SSN and VS, were played at a combined level of 60 dB SPL. Individual, non-ear-specific linear amplification was applied to the stimuli played to the HI listeners, following the NAL-RP prescription formula (Dillon, 2012, chapter 10), to compensate partly for their hearing loss. The spatially separated configuration was designed by playing the left noise only through the left channel of the headphones and the right noise only through the right channel of the headphones, thus providing infinite broadband ILD.

Figure IV-2 shows the mean measured SRTs as a function of the masker type (circles), whereby the black and grey colours refer to the data collected with NH listeners and HI listeners, respectively. The open symbols represent the measured SRTs in the co-located configuration ("co-loc") and the filled symbols show the SRTs for the spatially separated configuration ("separ"). The downward triangles correspond to the proposed model predictions and follow the same pattern of color and filling as the data. The good correlation with the data and low prediction errors ($r = 0.98$; $r_s = 0.90$; $MeanErr = 1$ dB; $MaxErr = 2.1$ dB) demonstrate that the model accurately predicts the effects across conditions and listener groups. In contrast, the Lav18 models are not able to predict well the difference between HI and NH listeners. The difference in measured SRTs across conditions between groups is about 7.2 dB, while the Lav18 models predict only 3.3 dB (and the proposed model predicts 6.4 dB). This discrepancy led also to worse model performances ($r = 0.89$; $r_s = 0.76$; $MeanErr = 2.1$ dB; $MaxErr = 4.0$ dB).

IV.3.1.2 Experiment of Rana and Buchholz (2018a)

Ten young NH listeners aged between 20 and 30 years (mean age of 23.2 years) participated in this experiment along with 10 older HI listeners aged between 57 and 78 years (mean age of 70.3 years; 4FAHL = 29.1 ± 8.0 dBHL). Only the non-stationary VS maskers were considered. The spatially separated configuration was simulated as in Rana and Buchholz (2016), involving infinite broadband ILD. Target and maskers were filtered to individually equalize audibility across frequency for each listener and then played at four different sensation levels (0, 10, 20 and 30 dB SL) relative to their individual SRT in quiet. The filters were designed using detection thresholds in quiet obtained with a SSN filtered into nine frequency regions. The measurement was done separately for each listener and frequency region. The gain corresponding to the detection thresholds determined the filters that were then applied to the stimuli played to the listeners. Only the broadband condition was considered here (lowband, midband and highband conditions were also tested in the original study). Amongst the 10 HI listeners, 1, 6 and 9 of them could not be tested at 10, 20 and 30 dB SL, respectively, due to loudness discomfort.

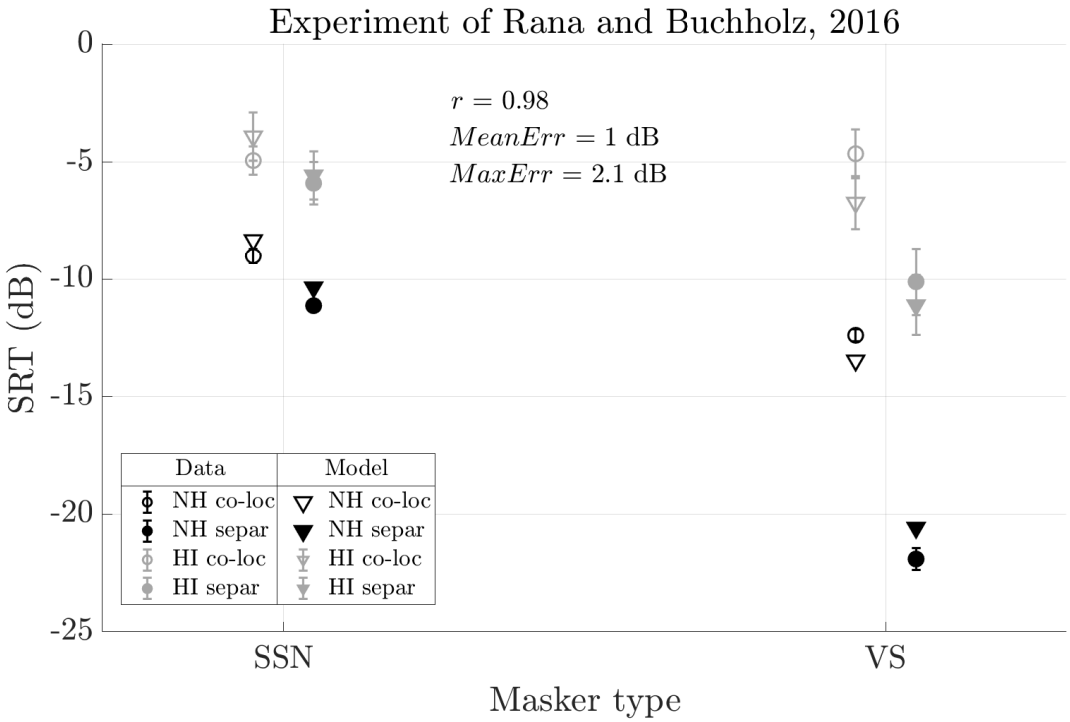


Figure IV-2: Mean SRTs with ±1 standard errors across NH listeners (blacks circles) and HI listeners (grey circles) measured by Rana and Buchholz (2016) as a function of masker type (SSN or VS). The two maskers were either co-located with the target in front of the listener ("co-loc", open symbols) or simulated on each side of the listener at ±90° ("separ", filled symbols). Mean predicted SRTs are displayed as downward triangles with the same filling and color patterns as the data.

Figure IV-3 presents the measured SRTs as a function of masker sensation level in two different panels, one for each group of listeners. The open circles correspond to the SRTs measured for the condition where the target and masker were co-located and the filled circles represent the SRTs obtained for the spatially separated configuration. The dashed and solid lines correspond to the model predictions for the co-located and spatially separated conditions, respectively. The proposed model predicts the data with a similar accuracy as for the experiment of Rana and Buchholz (2016) ($r = 0.98$; $r_s = 0.98$; $MeanErr = 1$ dB; $MaxErr = 3.1$ dB). The predictions provided by the Lav18 models are as good as the proposed model ($r = 0.97$; $r_s = 0.97$; $MeanErr = 1.4$ dB; $MaxErr = 3.3$ dB), even though a general underestimation of the SRTs for the HI listeners can be observed, as well as an overestimation of 2.5 dB for the NH listeners at 0 dB SL.

IV.3.1.3 Experiment of Rana and Buchholz (2018b)

Ten young NH participants aged between 25 and 41 years (mean age of 33.5 years) and 13 older HI listeners aged between 69 to 79 (mean age of 74 years; 4FAHL = 31 ± 8 dB HL) were involved in this experiment. Only the non-stationary VS maskers were tested and played at 60 dB SPL. Different amplification strategies were applied to the stimuli for the HI listeners to compensate partly for their hearing loss, but only the individual, non-ear-specific linear (NAL-RP) amplification was considered here. Three different spatially separated configurations were tested in addition to a co-located configuration. The first one was spatialized using natural ILD and ITD, the second one involved natural ILD but no ITD, and the last one applied the same process as the two previous experiments with infinite broadband ILD. The three spatially separated configurations are in the following referred to as "Natural", "ILD", "Infinite ILD".

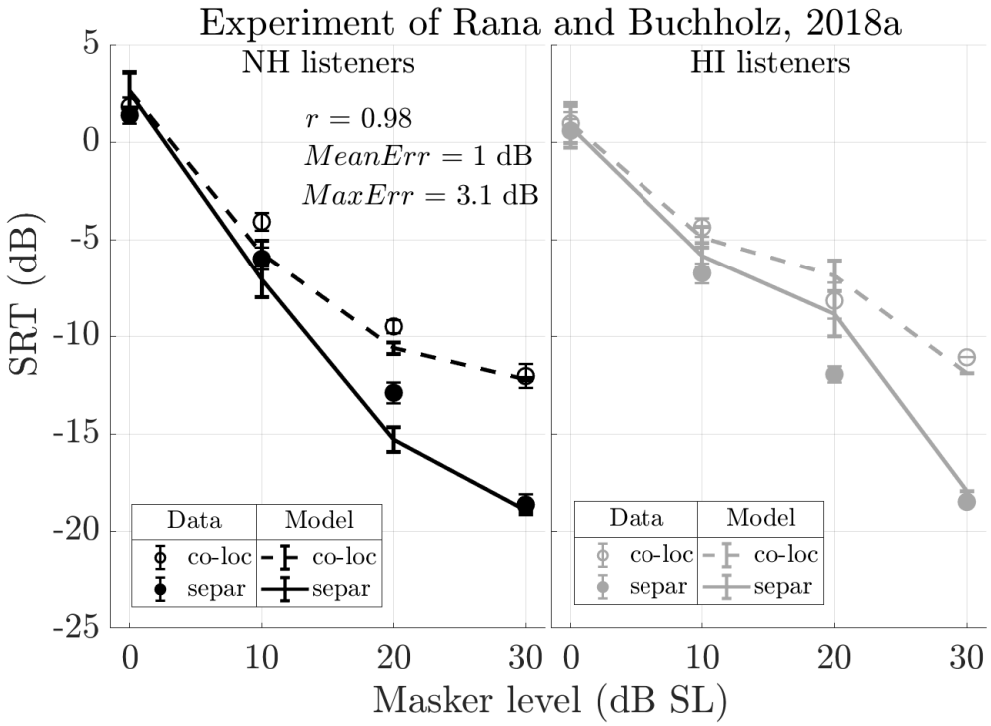


FIGURE IV-3: Mean SRTs (circles) with ±1 standard errors across NH listeners (left panel) and HI listeners (right panel) measured by Rana and Buchholz (2018a) at four overall masker sensation levels. Predicted SRTs are plotted with lines. The two maskers were VSs either co-located with the target in front of the listener ("co-loc", open symbols for data and dashed lines for predictions) or simulated on each side of the listener at ±90° ("separ", filled symbols for data and solid lines and predictions).

Figure IV-4 presents the measured SRTs (filled circles) as a function of spatialization method. The black and grey symbols correspond to the mean measured SRTs of the NH listeners and HI listeners, respectively. The downward triangles show the proposed model predictions; using the same color pattern as the data. The same model accuracy as for the two previous experiments (Rana and Buchholz, 2016, 2018a) is obtained ($r = 0.97$; $r_s = 0.95$; $MeanErr = 1.1$ dB; $MaxErr = 2.4$ dB). Concerning the Lav18 models, the difference in predicted SRTs across spatial conditions between the HI and NH listeners is about 0.3 dB, which is about 5 dB less than what is observed in the data. This discrepancy led to a lower correlation and higher errors ($r = 0.71$; $r_s = 0.79$; $MeanErr = 2.7$ dB; $MaxErr = 3.6$ dB) than those obtained with the proposed model.

IV.3.1.4 Individual differences between listeners

In order to investigate the accuracy of the model in predicting individual differences, a correlation analysis was conducted to compare the predicted SRTs with the measured SRTs for the HI listeners. For reference purposes, the correlation between the 4FAHL and the measured SRTs was also calculated. Only, the experiment involving the most HI listeners is considered here as an example, i.e., the one from Rana and Buchholz (2018b). The results for each spatial conditions are shown in Fig. IV-5 and results across spatial conditions are shown in Fig. IV-6. The 4 correlation coefficients between the 4FAHL and measured SRTs ($r = 0.92, 0.88, 0.82, 0.82$) within each spatial condition (Co-located, Infinite ILD, Natural and ILD) are higher than the coefficients between the predicted and measured SRTs ($r = 0.84, 0.84, 0.79, 0.78$). However, the correlation between all predicted and measured SRTs across spatial conditions ($r = 0.88$)

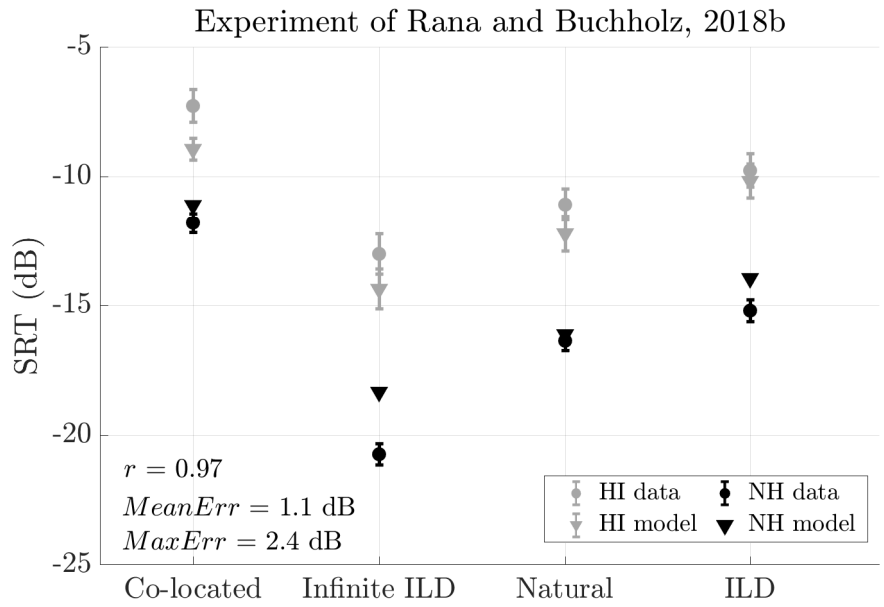


Figure IV-4: Mean SRTs with ±1 standard errors across NH listeners (in blacks circles) and HI listeners (in grey circles) measured by Rana and Buchholz (2018b) as a function of spatialization method. The two maskers were VSs. Predicted SRTs are displayed as downward triangles with the same color pattern as the data.

is higher than the correlation between 4FAHL and measured SRTs ($r = 0.65$). Hence, the 4FAHL explains slightly better the measured SRTs for a given spatial condition, but only the model is additionally able to describe the difference in SRTs across spatial conditions, which is not surprising as the 4FAHL was not designed to predict SRM and is independent of spatial condition.

IV.3.2 Dataset involving only NH listeners

In order to verify the backward compatibility of the proposed model, predictions for two experiments with only NH listeners are compared to previous model versions (Lav12 and Col13 models; see Sec. IV.2.3). As done above (and for the same reason), a NH listener was simulated here with 0 dB HL at all frequencies for both ears.

IV.3.2.1 Experiment 1 of Lavandier et al. (2012)

A meeting room was simulated through headphones, where a target was presented at 0.65 m and 25° of azimuth from the listener simultaneously with a SSN. The SSN was placed at one of two tested distances (0.65 m and 5 m, referred to as "Near" and "Far") and one of the three tested azimuths (-25°, 0°, 25° referred to as "Left", "Front", "Right"). Anechoic target and masker signals were convolved with BRIRs recorded at each tested position. Spectral-envelope impulse responses (SEIRs) were also tested. They were short binaural impulse responses artificially obtained by removing the reverberation tail and ITD of the BRIRs, while preserving their spectral envelope at each ear (and the resulting ILD).

Figure IV-7 presents the measured SRTs as a function of masker position. The black circles and grey squares show the SRTs measured with the BRIRs and SEIRs, respectively. The solid lines correspond to the predictions of the proposed model, while dashed lines present the predictions of the Lav12 model. The predictions of both models are very similar, the observed difference in each condition never exceeds 0.5 dB. The proposed model performance statistics ($r = 0.97$; $r_s = 0.95$; $MeanErr = 0.4$ dB; $MaxErr = 0.6$ dB) are therefore also very similar to

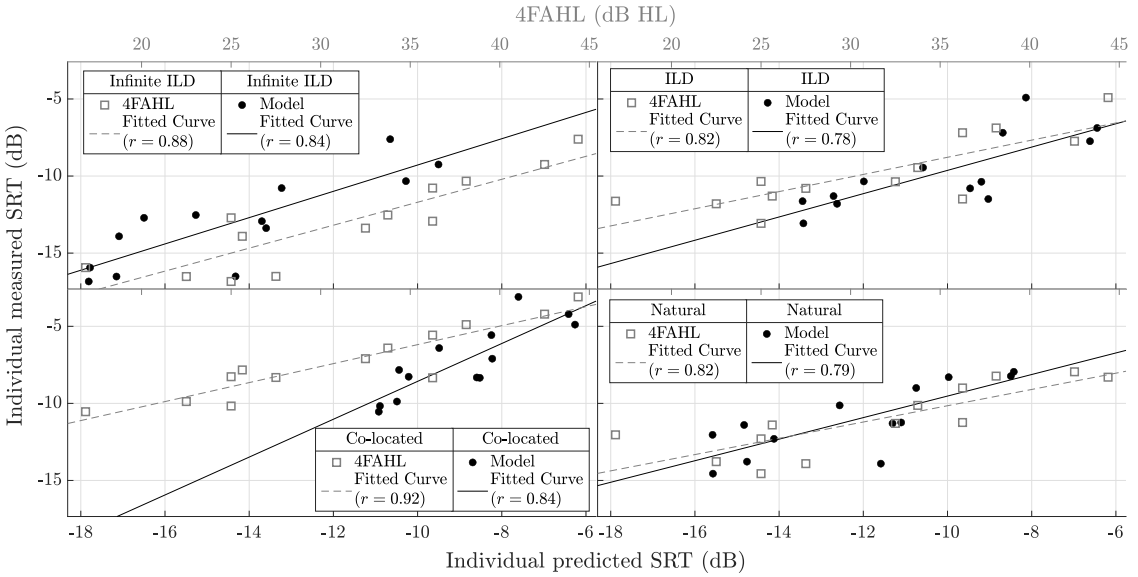


FIGURE IV-5: Individualized measured SRTs for HI listeners as a function of 4FAHL (grey squares) or individualized predicted SRTs (black circles). Only the experiment from Rana and Buchholz (2018b) is considered, each panel represents a spatialization method. The dashed and solid lines show the linear regressions between individualized measured SRTs and 4FAHL or individualized predicted SRTs, respectively.

the performances of the Lav12 model ($r = 0.98$; $r_s = 0.93$; $MeanErr = 0.3$ dB; $MaxErr = 0.6$ dB).

IV.3.2.2 Experiment 2 of Collin and Lavandier (2013)

Target and noise signals were presented diotically. The noise was a SSN either unmodulated or modulated with the broadband temporal envelope of 1, 2 or 4 simultaneous voices (the modulated noise changed from one target sentence to another while adaptively measuring the SRT, "variable" conditions, Collin and Lavandier, 2013).

Figure IV-8 presents the measured SRTs as a function of masker modulation plotted as circles. The upward triangles display the predictions for the proposed model and the downward triangles correspond to the Col13 model. The predictions of the two models are almost identical with a maximal difference of 0.2 dB. The models also predict the data very well with identical performance statistics ($r = 0.93$; $r_s = 1$; $MeanErr = 0.5$ dB; $MaxErr = 0.8$ dB).

IV.4 Discussion

IV.4.1 Improvements obtained with the new model

The proposed model has been optimized on three datasets involving NH and HI listeners. It accurately describes the effects of hearing loss and presentation level on speech intelligibility for NH and HI listeners using a single model version. This is a clear improvement compared to the preceding models of Lavandier et al. (2018), which required different parameter values to define the internal noise for NH and HI listeners, thus effectively resulting in two different model versions. The improvement is achieved here by the new implementation of the internal noise that now depends on the external sound level and divides the listener's hearing loss into a proportion η and $1 - \eta$ that roughly reflects the different effects of IHC and OHC loss on speech intelligibility. The proposed model also predicts the two NH datasets as accurately as the previous models it is based on, which validates its backward compatibility. This highlights that the implementation of the internal noise considerably extends the scope of application of the

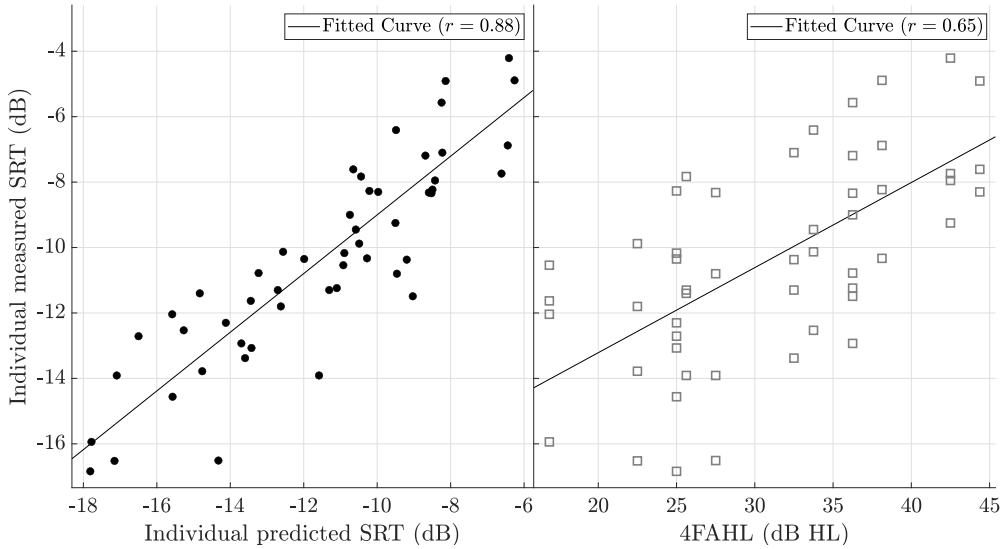


Figure IV-6: Individualized measured SRTs for HI listeners as a function of 4FAHL (right panel) or individualized predicted SRTs (left panel) across spatialization methods involved in Rana and Buchholz (2018b). The solid line within each panel shows the linear regression.

model by taking hearing loss into account, without compromizing its previously demonstrated accuracy. Note that the model still needs to be tested on data not used to define its parameters, so that its predictive power can be further evaluated.

The performance statistics obtained with the Lav18 models are similar or worse than the proposed model, namely over the three experiments $r \geqslant 0.71$, $r_s \geqslant 0.76$, $MeanErr \leqslant 2.7$ dB and $MaxErr = 4$ dB; as opposed to $r \geqslant 0.97$, $r_s \geqslant 0.90$, $MeanErr \leqslant 1.1$ dB and $MaxErr = 3.1$ dB (see Appendix B). The main reason is the definition of the internal noise by Lavandier et al. (2018) that leads to two main issues: (1) an underestimation of the effect of hearing loss on better-ear listening for the experiment of Rana and Buchholz (2016) and Rana and Buchholz (2018b) at least at a noise level of 60 dB SPL; (2) an overprediction of the effect of audibility close to the hearing thresholds (Rana and Buchholz, 2018a). In other words, the issue (1) likely means that the level of the internal noise is not high enough to limit the better-ear SNR for HI listeners. Making this level dependent on the external sound level solves this issue. The issue (2) is solved by splitting the audiograms into proportions that resembles OHC and IHC loss, which affects both the predictions at low sensation levels and the effect of increasing audibility. At low sensation level, when the internal noise level is equivalent to $\eta\, T(n) + B$, the difference between NH and HI listeners is reduced because only 0.7 of the pure-tone audiogram is taken into account. This decreases substantially more the level of the internal noise of HI listeners compared to NH listeners (because of the higher hearing thresholds of HI listeners), so that the predictions are similar at low sensation levels.

The model performance statistics when predicting the experiment of Rana and Buchholz (2018a) are similar. This is due to the parameter that defines the broadband level of the internal noise, which in Lavandier et al. (2018) was optimized separately for each listener group, to predict the difference between conditions within groups and not the difference between groups. Here, the Lav18 models successfully predict the data from Rana and Buchholz (2018a) since there is no significant difference in SRTs between groups. However, the predictions are worse for the experiments from Rana and Buchholz (2016, 2018b) because there is a difference in SRTs between both groups that the Lav18 models cannot predict.

IV.4.2 Further exploration of the internal noise implementation

Here, a more detailed analysis is provided to illustrate the effect of the external noise level and hearing loss on the internal noise described in Eq. IV.1, and the expected impact on speech

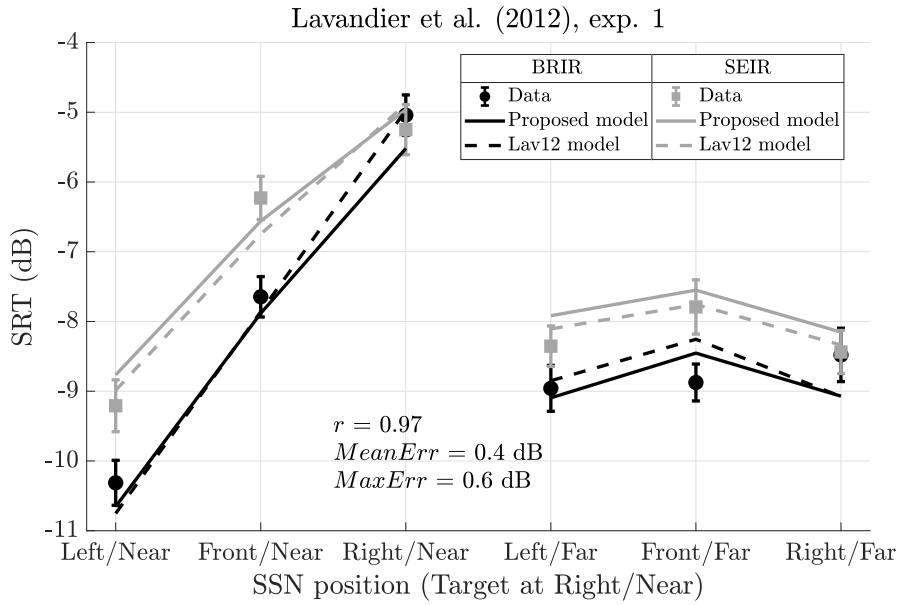


FIGURE IV-7: Mean SRTs with ± 1 standard errors across listeners as a function of SSN position, measured by Lavandier et al. (2012). The conditions spatialized with BRIRs are plotted as black circles and those with SEIRs as grey squares. Predicted SRTs from the proposed model are displayed as solid lines and those predicted with the Lav12 model are plotted as dashed lines, with the same color pattern as the data. Model performance statistics are displayed only for the proposed model.

intelligibility. For this analysis, the average of all the HI listeners' audiograms across the three main experiments is applied as an example hearing loss. The values after being rounded to the closest multiple of 5 are 15, 20, 25, 35, 45, 50, 60, 60 dB HL for the frequencies 250, 500, 1000, 2000, 3000, 4000, 6000, 8000 Hz.

Figure IV-9 shows the internal noise spectra for NH and the example HI listener at three different external noise levels (40, 60, and 80 dB SPL). It can be seen that the spectral shape of the internal noise for the NH listener is independent of the external noise level and solely determined by their NH audiogram in dB SPL (i.e., 0 dB HL). The overall internal noise level is identical for the two lowest external noise levels, but then increases substantially at the highest level due to the emerging contribution of the IHC loss component described by Eq. IV.3. The same behaviour can be observed for the HI listener, except that due to the sloping, high-frequency hearing loss the overall internal noise level is much higher than for the NH listener and increases with increasing frequency. Note, that the effect of hearing aid amplification was disregarded here, which would increase the external masker level and thus the internal noise level. Considering the average hearing loss and the masker stimuli used in the three main experiments, linear amplification according to NAL-RP would result in an increase of the external noise (broadband) level of about 5 dB.

Figure IV-10 shows the internal noise level for a NH and the example HI listener (see above) as a function of external noise level for the model center frequencies of 516, 2017 and 3937 Hz. For all frequencies and both listeners, the internal noise level is constant at low external noise levels and solely determined by the OHC loss component described by Eq. IV.2. Above a certain external noise level, the internal noise level starts to increase, due to the emerging contribution of the IHC loss component of Eq. IV.1. Further increasing the external noise level then leads into a dB-for-dB increase of the internal noise level, as is described by Eq. IV.3. Again, the overall internal noise level is much higher for the example HI listener than the NH listener, and due to the sloping, high-frequency hearing loss, the influence of the IHC loss component also starts at lower external noise levels than for NH listeners, which becomes further pronounced with increasing frequency. The influence of the NAL-RP amplification on the internal noise levels

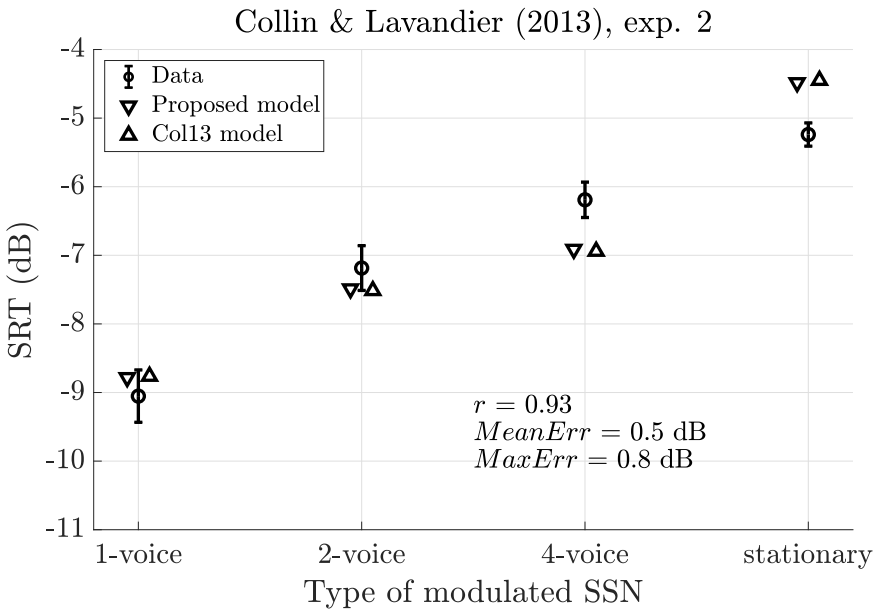


Figure IV-8: Mean SRTs with ± 1 standard errors across listeners as a function of masker modulation depth, measured by Collin and Lavandier (2013) in their "variable" masker condition (circles). Target and noise signals were presented diotically. The SRTs predicted with the proposed model are displayed as downward triangles and those predicted with the Col13 model as upward triangles. Model performance statistics are displayed only for the proposed model.

is plotted as dashed lines. The behaviour of the internal noise levels is similar but the curves are shifted to the left along the x-axis by about 5 dB, which represents the broadband increase induced by amplification. Hence, the pure-tone audiograms increase twice the internal noise levels when the signal is amplified according to the hearing loss (as the NAL-RP amplification does): "directly" with the terms $(1-\eta)T(n)$ and $\eta T(n)$ as well as "indirectly" by increasing the external noise level N.

For the noises applied in the main experiments with a broadband level of 65 dB SPL, which represents the level of a moderately noisy environment (Weisser and Buchholz, 2019), the level in the model filter bands at the three considered frequencies is 53, 35, and 43 dB SPL. Given that for NH listeners the corresponding internal noise levels are all below 6 dB SPL, the effect of the internal noise on predicted speech intelligibility is negligible. With respect to the present study, the effect of the internal noise is only relevant in the experiment of Rana and Buchholz (2018a) at the softest sensation level of 0 dB SL. This is already different for the example HI listener, who has a mild-to-moderate hearing loss and, at an external noise level of 65 dB SPL, involves internal noise levels of 17, 34, and 46 dB SPL. Considering the level fluctuations in the masker, the internal noise will swamp (or dominate) a significant portion of the masker signal at mid and high frequencies.

IV.4.3 Is the proposed internal noise concept in line with the literature ?

The proposed implementation of the internal noise shares common aspects with other internal noises described in the literature, in particular the one implemented in the monaural SII (ANSI S3.5, 1997). The SII is a SNR-based index (between 0 and 1) that predicts speech intelligibility in noise. It is obtained by integrating the SNRs across frequency bands using a SII weighting as well as a distortion factor that degrades intelligibility when target levels are higher than a reference. The internal noise of the SII is simply realized by the listener's pure-tone audiograms averaged across ears and added to a reference internal noise spectrum level (the internal noise level for a NH listener with 0 dB HL). However, there is no distinction of the influences of IHC and OHC loss and the external noise level does not affect the internal noise. The similarity between the two models comes from the way the highest level between the internal noise and

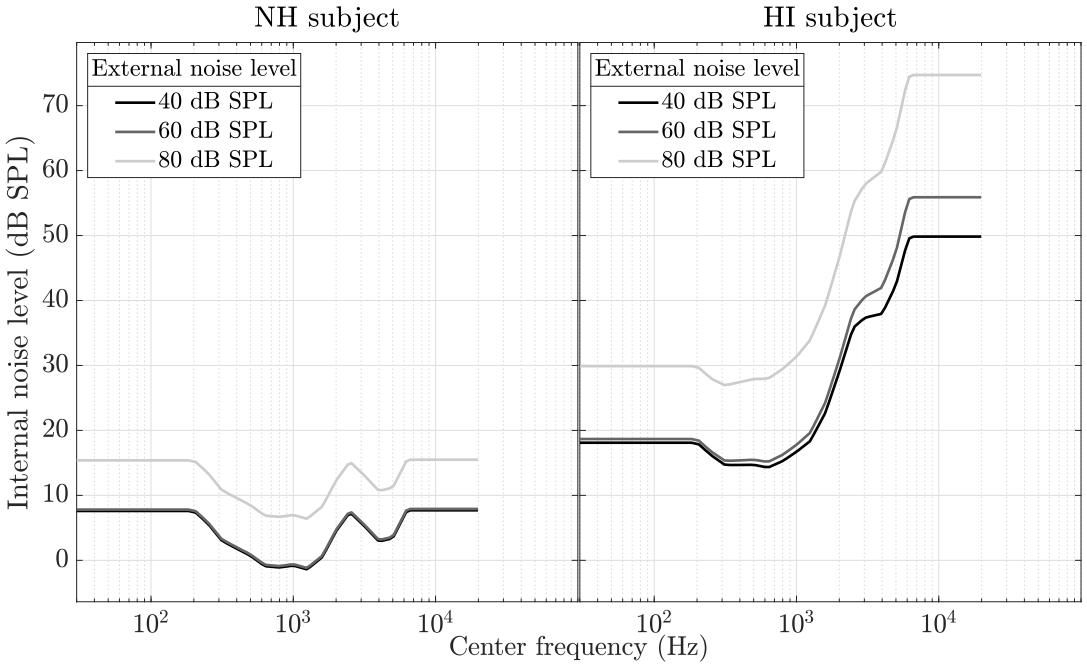


FIGURE IV-9: Internal noise level as a function of the model center frequency based on either an audiogram with 0 dB HL (NH listener, left panel) or the audiogram averaged across all HI listeners considered in this study (HI listener, right panel). In each panel, the three internal noise spectra are plotted as solid lines and shade of grey for an external noise level of 40, 60 and 80 dB SPL.

the external noise is chosen to compute the SNR in each frequency band. The influence of the distortion factor in the SII could be compared to the influence of the external-level-dependent component of the internal noise in our model, because both degrade speech intelligibility when the external signal level is higher than a reference. However, they differ in two major ways. First, our internal noise relies on the broadband level of the external noise, whereas the SII distortion factor relies on the level of the target speech in each frequency band. Furthermore, the SII distortion factor is applied to the SNR *a posteriori*; while our level-dependent internal noise is directly involved in the computation of the SNR.

Ching et al. (1998) proposed to modify the SII in order to account for the deficit of speech understanding in HI listeners at high sound levels that cannot be explained by loss of audibility. They tested nine procedures to conclude that "the sensation level, as well as the sound presentation level, affect a listener's ability to make use of audible information". The internal noise proposed here takes into account these two characteristics since the listener's audiogram and the external noise level are included in the SNR computation.

Plomp (1978) proposed a formula to predict SRTs as a function of the listener's hearing loss and external noise level. The listener's hearing loss is considered in two components (see Sec. IV.1 for more details), which can be compared to the current rough estimate on IHC and OHC losses. However, Plomp's formula does not consider the listener hearing loss as a function of frequency and it is quiet unclear how the formula account for the effect of binaural listening. Hence, the scope of application of the current model is broader than Plomp's one.

The binaural speech intelligibility "EC/SII" model developed by Beutelmann and Brand (2006) and its extension "BSIM" proposed by Beutelmann et al. (2010) combines an EC stage with the monaural SNRs at both ears to predict intelligibility. Both models implement internal noises defining their spectrum levels by adding a parameter equal to 4 and 1 dB,[12] respectively, to a spectrum shaped to the listener's audiograms. In either model, the internal noises impact the

[12]The value of 4 dB in the version of Beutelmann and Brand (2006) was taken from the literature, while the value of 1 dB was chosen by Beutelmann et al. (2010) because it provided the best correlation between data and predictions in a reference experiment.

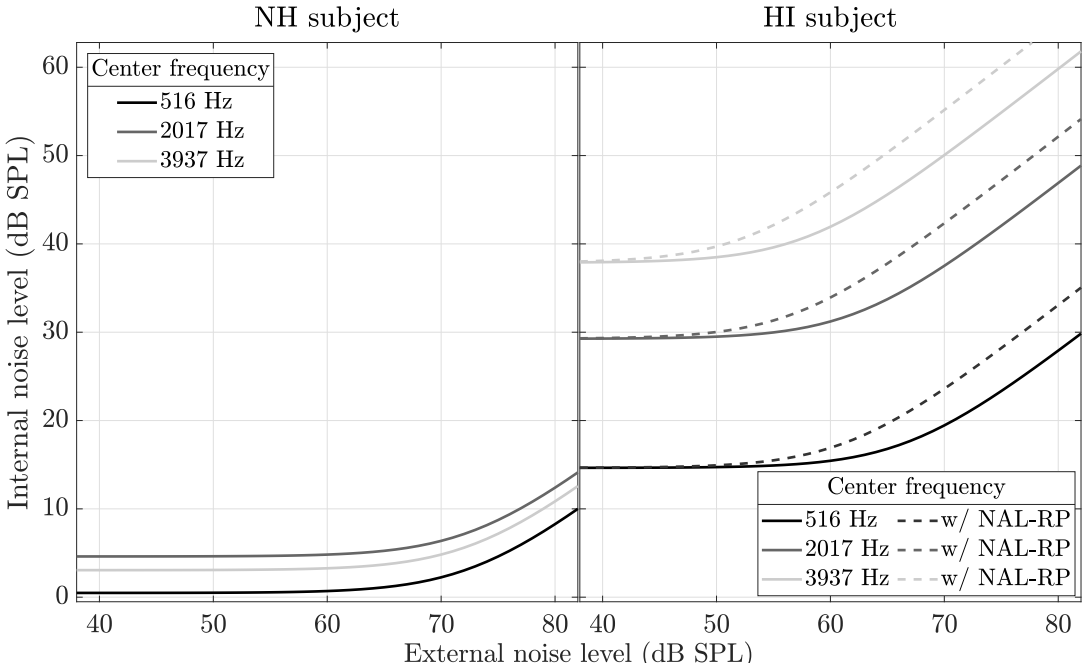


Figure IV-10: Internal noise levels at 3 given center frequencies of the model as a function of the external noise level. In the left panel, the internal noise level is generated using an audiogram with 0 dB HL at any frequency and ear (NH listener). In the right panel, the audiogram averaged across all HI listeners considered in this study (HI listener) is used to compute the internal noise level. In the same panel, the influence of the NAL-RP amplification on the internal noise level is plotted as dashed lines.

model EC-stage, and thus binaural unmasking. The internal noise can influence the coherence of the resulting overall (internal + external) noises at the listener's ears, thus reducing the efficiency of the EC mechanism. However, the internal noises also impact the monaural paths of the model, because they tend to reduce the monaural SNR in each frequency band. The internal noise implementations in both models do not take into account the external noise level nor do they differentiate between estimated contributions of IHC and OHC loss. In that way, their implementations are closer to the one of Lavandier et al. (2018), in which the audiogram is used to spectrally shape the internal noise and a frequency-independent parameter is applied to define its broadband level.

Our current internal noise implementation can decrease the SNR at the better ear, but it never reduces the overall masker coherence. The impact of the internal noise (including hearing loss) on binaural unmasking is currently modelled such that binaural unmasking is not affected at all, until one of the external signals (masker or target) gets quieter than the internal noise, in which case the binaural unmasking advantage is set to zero in the corresponding frequency band and time frame. This very crude implementation might need refinement in the future, and will require testing on data that specifically assess binaural unmasking for NH and HI listeners.

The definitions of the two internal noise components proposed by Bernstein and Trahiotis (2008) and presented in Sec. IV.1 match closely the asymptotic behavior of our internal noise: at low external stimulus level, the internal noise level is equal to $\eta T(n) + B$ (see Eq. IV.2), which can be considered as an internal noise floor that is independent of the external stimulus. For a higher external stimulus level, the internal noise increases with the external noise level following a dB-for-dB rule (see Eq. IV.3).

IV.4.4 Hypotheses and limitations of the proposed model

Regarding the individual predictions shown in Sec. IV.3.1.4, it is not surprising that the model cannot fully explain the variance seen in the individual SRTs, because it does not consider

all aspects of hearing loss. For instance, even though the model takes into account SRM, it is only based on the audiograms and does not consider other aspects of hearing loss such as auditory filter broadening (Glasberg and Moore, 1986) or loss of ITD sensitivity (King et al., 2014; Neher et al., 2011). Moreover, some variation in the data might be influenced by cognitive effects (Neher et al., 2012), motivation or speech material effects (e.g., differences in intrinsic intelligibility of the sentence lists), that are not considered in the model. Given that the NH listeners who participated in all considered experiments were much younger than the HI listeners, it is also likely that age may have contributed to some of the differences seen between groups (Füllgrabe et al., 2015; Neher et al., 2012; Schneider and Pichora-Fuller, 2001), and thus, to the differences seen between data and model predictions. However, since hearing loss naturally increases with age, it is difficult to disentangle the effects of aging and hearing loss.

With reference to the experiments from Rana and Buchholz (2016, 2018a), the proposed model tends to slightly underestimate the difference in SRTs between NH and HI listeners when VS maskers are applied, which indicates that the model slightly underestimates the advantage associated with modulations in the masker envelope for the NH listeners and/or overestimates this advantage for the HI listeners by about 2 dB. This might be explained by the fact that the proposed model does not take into account the larger gap detection thresholds (or reduced temporal resolution) that are typically observed in HI listeners (e.g., Fitzgibbons and Wightman, 1982), which can degrade the benefit associated with listening in the masker's gaps.

Even though the current model can successfully describe the data from three experiments, its binaural unmasking component was tested only in the "Natural" configuration of Rana and Buchholz (2018b). Existing studies have shown that an increase of hearing loss leads to lower ITD sensitivity (e.g., Füllgrabe and Moore, 2017; Santurette and Dau, 2012). Furthermore, a significant negative correlation between ITD sensitivity and SRT was reported by Strelcyk and Dau (2009) and Neher et al. (2011), but in only three test conditions, which might not be enough to draw general conclusions. Our current implementation of the effect of hearing loss on binaural unmasking is very simplistic, which solely considers the effect of reduced audibility but not the effect of reduced ITD sensitivity. This might prove insufficient to fully predict binaural unmasking in HI listeners when more relevant conditions are tested. In addition, as discussed in Sec. IV.4.3, the internal noise implemented in the binaural model proposed by Beutelmann and Brand (2006) or Beutelmann et al. (2010) and the results of Bernstein and Trahiotis (2008) suggest that the coherence of the signal resulting from the combination of the external and internal noises could be lower than the external noise coherence, thus impacting the efficiency of binaural unmasking. Hence, the proposed model should be further tested on datasets that are more relevant for binaural unmasking, including differences in ITDs between the target and masker signals as well as varying the masker coherence at the listener's ears (Lavandier and Culling, 2010), but measured for HI listeners.

Future studies should also test the proposed model on additional acoustic conditions (such as Beutelmann et al., 2010, did) to further evaluate its general applicability to predict speech intelligibility in HI listeners. For instance, only energetic noise maskers have been tested so far because the model cannot take into account the informational masking that can occur with speech maskers (e.g., Kidd and Colburn, 2017). Furthermore, it would be relevant to consider reverberant conditions, because reverberation degrades speech intelligibility in multiple ways (Lavandier and Culling, 2008). It impairs better-ear listening by filling the masker's gaps (Collin and Lavandier, 2013), and binaural unmasking by decreasing the interaural coherence of the masker (Lavandier et al., 2012). These two effects were well predicted by the NH version of the model, but they were not specifically tested for HI listeners. At high levels, reverberation can also be detrimental to the intrinsic intelligibility of the target, even in the absence of a masker. This effect was successfully predicted by a modified version of the NH model (Leclère et al., 2015), but again, it remains to be tested how well combining this previous model with our internal noise implementation predicts the effects of reverberation for HI listeners. Finally, the non-linear processing present in hearing aids, such as spectral subtraction or non-linear amplification, have not been involved in the present study. The model will need to take into account the influence of this processing to predict intelligibility for HI listeners wearing their hearing aids. In this regard, the general effect of amplification on the internal noise of the model needs to be investigated. Currently, amplification does not only increase audibility of the input signals, but since it increases the overall masker level it also increases the internal noise level of the IHC loss-related component (see Sec. IV.4.2). Hence, testing the model on a

dataset involving conditions with and without amplification would be informative on the effect of amplification on internal noise level.

With respect to the internal noise implementation of the proposed model, the underlying assumptions described in Sec. IV.2.2 may need to be revised in the future. The first assumption was that the broadband SNR is below 0 dB, so that the external level can be approximated by the masker level to define the internal noise. However, for the experiment of Rana and Buchholz (2018a), the SNRs for the 0 dB SL condition are above 0 dB. In that case the assumption is violated, even though the average target level was probably not high enough to affect significantly the level of the listener's internal noise, so that in practice, this violation did not impair the predictions. In addition, as shown by Smeds et al. (2015), daily situations mostly involve positive SNRs, hence, this assumption must be reconsidered to predict intelligibility in real life situations. Similarly, the assumption that the internal noise is solely based on the broadband level of the external noise, may need to be revised. The distortion factor applied in the SII to limit speech intelligibility at high speech levels, for example, does not only depend on the broadband level of the target speech, but also on its spectrum. Within the proposed model, such frequency dependence of the internal noise might need to be considered in the future.

Another assumption concerns the parameter η (equal to 0.7) that is used to divide the listener's hearing loss into proportions η and 1-η, interpreted as a rough estimate of OHC and IHC loss (Sec. IV.2.2). Even though this value is close to the value used by Bruce et al. (2013) and Scheidiger et al. (2018), who attribute two thirds of the hearing loss to OHC loss to predict monaural speech intelligibility, other studies have shown that the proportion of IHC and OHC loss varies across listeners and place (or frequency) on the Basilar membrane (e.g., Moore and Glasberg, 2004; Pieper et al., 2018). In this regard, the proposed implementation of a single-value η was an intentional simplification, with the aim of limiting the number of fitting parameters and the complexity of the model. However, this may need to be revised in the future.

Finally, the model was both optimized and evaluated on the same dataset involving NH listeners and HI listeners with mild to moderate-severe hearing loss. Even though the model has only a small number of fitting parameters when compared to the number of tested data points, it still needs to be verified on data that is not used to define its parameters. Moreover, the model needs to be tested for more severely impaired HI listeners and listeners with asymmetric hearing loss to evaluate its relevance to predict speech intelligibility for arbitrary listeners.

IV.5 Summary

A binaural model is proposed that uses the listener's audiogram to predict the effects of hearing loss and presentation level on speech intelligibility in noise for HI and NH listeners. This was done by splitting the audiogram into proportions interpreted as rough estimates of OHC and IHC loss and highlighting that the internal noise consists of two components, one related to elevated thresholds and the other considering supra-threshold effects that depend on the external noise level. The resulting model shows similar predictions to its previous model versions when considering data measured only for NH listeners, and provides accurate results when predicting datasets involving NH and HI listeners on which it has been optimized. These involve and experimental designs that aimed to evaluate the effects of audibility, spatial configuration and noise types on speech intelligibility. Across the 5 experiments considered in the study, the model predictions are accurate as quantified by r greater or equal to 0.93, $MeanErr$ not exceeding 1.1 dB and $MaxErr$ equal to 3.1 dB. The influence of hearing loss on binaural unmasking needs to be further investigated, which is the goal of the next chapter.

V

The contribution of binaural unmasking to speech intelligibility in noise for normal-hearing and hearing-impaired listeners

This chapter presents a study that considered a binaural speech intelligibility model and an experimental approach to investigate the relationship between hearing loss and binaural unmasking for speech intelligibility in noise. Ethical clearance was received from the Macquarie University Human Sciences Ethics Committee (approval number: 5201955629923, see Appendix D). Sensitivity to ITD and SRTs were measured for NH and HI listeners in the experimental study. The model developed in the previous chapter (Vicente et al., 2020) was used to model the intelligibility data. It successfully predicted the SRTs, except that a modification of the formula computing the binaural unmasking advantage in the model was required in order to account for the effect of presentation level. The content of this chapter is under review to be published in the Journal of the Acoustical Society of America.

V.1 Introduction

Binaural unmasking has been well studied and documented for NH listeners. The binaural unmasking advantage has been often measured as the difference in SNR required to detect a tone in noise between a diotic condition and a dichotic condition with at least one signal, target or masker, presenting an ITD or an IPD. Durlach (1972) provided a review of some factors influencing binaural unmasking. The binaural unmasking advantage increases up to about 15 dB with the difference in IPD (which can be associated with a difference in source azimuth for real-life sources) between tone and noise, as well as with noise coherence and with stimulus presentation level. The results of Durlach (1972) and Zwicker and Zwicker (1984) show that the binaural unmasking advantage increases with sensation level when it is varied from 0 to 50 dB SL, and then reaches a plateau. Binaural unmasking has been also studied in the context of speech intelligibility in noise, by measuring differences in SRT relative to a given reference. The reference condition is a condition presenting the lowest difference in ITD between target and masker signal (e.g., when they are spatially co-located or aligned in azimuth). The binaural unmasking advantages are around 4-6 dB across experiments involving this paradigm (Bronkhorst and Plomp, 1988; Goverts and Houtgast, 2010; George et al., 2012; Lavandier and Culling, 2010). The factors influencing the contribution of binaural unmasking to speech intelligibility in noise are the same as those reported before for tone detection in noise. In addition, George et al. (2012) investigated the effect of temporal fluctuations in the masker envelope and found that the binaural unmasking advantage for speech was lower than for an unmodulated noise.

Durlach et al. (1981) reviewed studies investigating the effect of hearing loss on binaural unmasking and concluded that hearing loss did not necessarily degrade binaural unmasking. In many cases, symmetrically HI listeners showed similar results to NH listeners. Listeners with asymmetric hearing loss generally performed worse, but mostly because signals were presented at equal level at the listener's ears rather than at equal sensation level. The results of studies that followed this review agreed with those general observations (e.g., Bronkhorst and Plomp, 1989; Goverts and Houtgast, 2010). For the few cases where HI listeners showed lower binaural unmasking advantages, Goverts and Houtgast (2010) found that supra-threshold deficits were the reason of this deterioration, in particular, the loss of signal phase and time coding.

In contrast to these masking studies, various studies have demonstrated that HI listeners show more difficulty than NH listeners in detecting an IPD (Füllgrabe and Moore, 2017; King et al., 2014; Neher et al., 2011). Normal-hearing listeners can detect an IPD for frequencies up to 1.4 kHz while HI listeners show lower sensitivity, i.e., a lower upper frequency for detection. For instance, on average across HI listeners, Füllgrabe and Moore (2017) measured an upper frequency of 699 Hz, while Santurette and Dau (2012) reported that most of their HI participants showed a threshold around 1 kHz. The correlation analysis of Santurette and Dau (2012) revealed a significant relation between the highest frequency at which an IPD could be detected and the binaural unmasking advantages measured for detection of a pure tone at 500 Hz and 1 kHz. However, they did not find any significant correlation between IPD detection and SRTs. This was different from Strelcyk and Dau (2009), Neher et al. (2011) and Neher et al. (2012), who did find a significant correlation for HI listeners between IPD detection thresholds and SRTs measured in a condition that maximized the difference in ITD between target and masker.

Culling et al. (2004, 2005) provided a BMLD formula based on the work of Durlach (1972) that predicts the binaural unmasking advantage for detection of tone in noise as well as for the SRT for speech in noise for NH listeners:

$$BU_{Adv} = 10log_{10}\left[\frac{k - cos(\theta_s - \theta_n)}{k - \rho}\right] \tag{V.1}$$

with

$$k = (1 + \sigma_\epsilon^2)exp(\omega_c^2\sigma_\delta^2) \tag{V.2}$$

where θ_s and θ_n are the IPDs of the speech and noise (in radians), ρ is the noise interaural coherence, ω_c is the center frequency of the given frequency band (in rad/s), and σ_ϵ and σ_δ are parameters with values equal to 0.25 and 0.000105 s (Durlach, 1972). The two parameters (also called "jitters") reflect the amplitude and time alignment errors that the auditory system makes during the equalization stage in the EC theory. Durlach (1972) considered these jitters as free parameters and evaluated their values by fitting data derived using tone detection in noise and chosen to avoid any audibility issue. Durlach (1972) was aware that the formula did not take into account the effect of presentation level or hearing loss, and until now the formula has not been modified to address these limitations.

The present study aimed to thoroughly investigate the effects of hearing loss, reverberation (that influences the noise interaural coherence), presentation level, masker envelope fluctuations and difference in azimuth between target and masker on the contribution of binaural unmasking to speech intelligibility in noise. The sensitivity of listeners in detecting an IPD was also measured in order to understand whether this was linked to a reduced binaural unmasking advantage. In addition, the binaural model developed by Vicente et al. (2020) for predicting the effect of audibility on speech intelligibility in envelope-modulated noises for HI and NH listeners was used to describe the effects observed in the data. The model consists of two components, one predicting the binaural unmasking advantage using the BMLD formula of Culling et al. (2004, 2005, Eq. V.1) and the other predicting the better-ear listening benefit (Lavandier and Culling, 2010). The current dataset was used to test the accuracy of the binaural unmasking component, i.e., to challenge the BMLD formula to predict data at low presentation levels as well as data collected with HI listeners.

V.2 General methods

Data was collected in two 1.5-hour sessions that were either separate or with a substantial break between them. The listeners received a hearing test measuring their pure-tone audiogram at the beginning of the experiment if their last audiogram measured in the laboratory dated back to a year or more. The listener's sensitivity for detecting an IPD was measured at the beginning of each session. SRTs were measured under 22 conditions (see Sec. V.2.3), randomly split over the two sessions. The experimenter was a native Australian-English research audiologist. Ethical clearance was received from the Macquarie University Human Sciences Ethics Committee (approval number: 5201955629923, see Appendix D).

V.2.1 Listeners

The 20 listeners involved in the study were native Australian-English speakers, 12 of them were NH listeners aged between 20 and 33 years (mean age of 25.6 years), and 8 were HI listeners aged between 19 and 80 years (mean age of 61.0 years). Listeners were considered NH if they had hearing thresholds below 20 dB HL from 250 Hz to 8000 Hz for both ears (one outlier allowed). The HI listeners had symmetric (less than 10 dB HL difference between ears from 250 Hz to 4000 Hz, one outlier allowed) mild-to-moderate sensorineural hearing loss. The 4FAHL and ± 1 standard deviation was 4.8 ± 2.4 dB HL for the NH listeners and 30.6 ± 5.8 dB HL for the HI listeners. Figure V-1 presents as grey solid lines the individual audiograms averaged across ears of the NH listeners (left panel) and HI listeners (right panel). The black solid lines show the mean audiograms across listeners. The listeners were paid for their participation, or the students attending courses at the University could receive course credit instead, in line with Macquarie University policy.

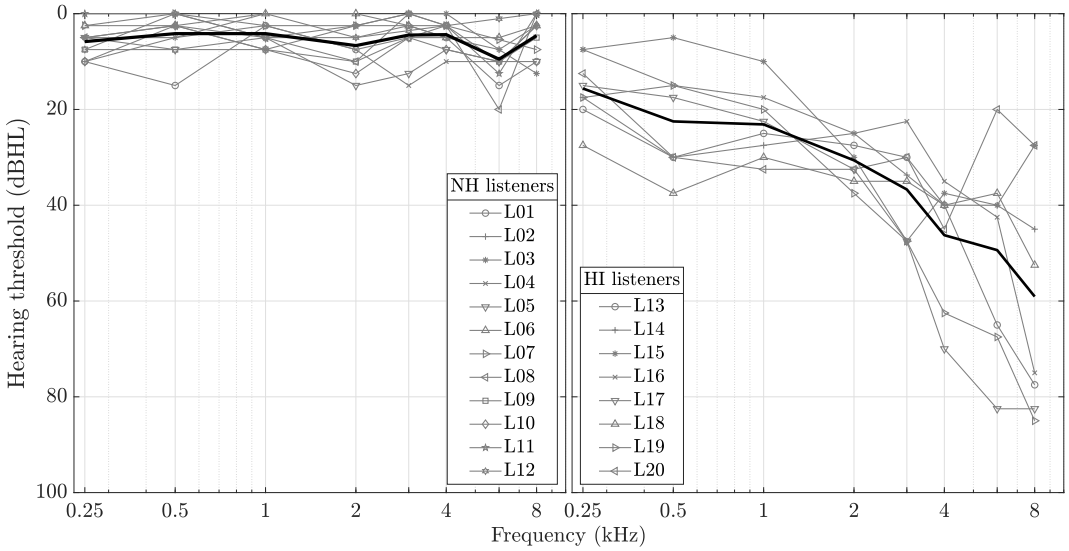


Figure V-1: Average audiograms across ears of the NH (left panel) and HI (right panel) listeners. Each listener is represented by a symbol and a single marker style. The black solid line represents the average audiogram across listeners in each panel.

V.2.2 IPD sensitivity measurement

The temporal fine structure adaptive frequency (TFS-AF) test developed by Füllgrabe et al. (2017) was used to measure sensitivity for detecting an IPD. It used a 2-interval 2-alternative forced-choice protocol. One interval consisted of a series of four 400-ms long tone bursts that were in phase at the ears and separated by 100 ms (including 20-ms raised-cosine rise/fall ramps). In the other interval, the second and fourth tones had an IPD equal to 180°. The listener's hearing thresholds at 0.25, 0.5, 1, 2 kHz were input to the software in order to play the tones at 30 dB SL to ensure good audibility. The intervals were played consecutively through headphones (Beyerdynamic DT990 Pro) in random order, with a gap of 500 ms between intervals. The listener, seated in front of a computer screen, had to choose which interval sounded more spatialized/diffuse by clicking on the button "1" or "2" displayed on the software interface. The frequency of the tones was varied adaptively following a 2-up, 1-down rule to measure the 71%-correct point on the psychometric function.

The test started with a training phase, in which the listener was asked to listen to a series of spatialized and non-spatialized tones at different frequencies. Then, the test began with tones at 200 Hz. After 8 reversals, the upper frequency limit for detecting a change in IPD (IPD_{freq}) was estimated as the geometric mean of the frequency values at the last 6 reversal points.

V.2.3 SRT measurements

Lavandier and Culling (2010) set up two experiments to systematically test the contribution of binaural unmasking to speech intelligibility in stationary noise for NH listeners. In their first experiment, they varied the difference in ITD of the target and masker by changing the target azimuth, while the influence of the masker coherence was tested by using different levels of reverberation in a simulated room. The idea in the present study was to extend this experiment to HI listeners and to use modulated noise in addition to stationary noise.

The target and masker sources were simulated in a room using binaural room impulse responses and the signals were played through headphones (Beyerdynamic DT990 Pro). In order to minimize the effect of better-ear listening, the listener's head was simulated as acoustically transparent using two virtual omnidirectional microphones located at the listener's earswith no virtual head between them. The room impulse responses were generated between each simulated ear and the target or masker positions. The stimuli were spatialized by convolving anechoic signals (see V.2.3.1) with the impulse responses of the room obtained with the simulation done by Lavandier and Culling (2010, experiment 1). The broadband sound level differences introduced by the spatialization process were removed by equalizing the broadband level at each ear to a reference level. This set the broadband ILD to 0 dB, although some frequency-dependent ILDs still occurred due to reverberation (room coloration).

The virtual room was simulated as a shoe box room with dimensions of 10 m x 6.4 m x 2.5 m (Length x Width x Height) and a single frequency-independent absorption coefficient (α) for all room surfaces, which was varied to control the level of reverberation.[13] The interaural masker coherence was varied by applying different level of room reverberation to the masker. The target was always anechoic, to avoid any effects of reverberation that were not relevant to test the contribution of binaural unmasking to speech intelligibility in noise (such as temporal smearing that degrades intrinsic speech intelligibility, e.g., Leclère et al., 2015).

V.2.3.1 Stimuli

The speech material was BKB-like sentences (Bench et al., 1979), which are short meaningful sentences of 4 to 7 words that can be understood by a typical 5-year old. The corpus contains 80 lists of 16 sentences that were read aloud by a female Australian-English speaker. The masker was either stationary SSN or VS. These noises were generated as described Rana and Buchholz (2016). The SSN was obtained by filtering a white noise to match the average spectrum of the entire corpus of BKB-like sentences. The VS was generated by decomposing separately two samples of running speech (taken from the LiSN-S speech corpus, Cameron and Dillon, 2007) in the spectral domain using the short-term Fourier transform with 75%-overlapping 20-ms-long Hanning windows. The magnitude of the resulting spectra was smoothed using the power spectrum of a Gammatone filter with a bandwidth of four ERB (Moore and Glasberg, 1983). The smoothed spectra were combined with a random phase and an inverse Fourier transform was applied for resynthesis, to get the signals back into the time domain. The two processed speech discourses were added, resulting in the VS masker that was significantly less intelligible than the target speech. Finally, the VS masker was filtered to match its long-term spectrum with the average spectrum of the entire corpus of BKB-like sentences.

The room was simulated with three levels of reverberation using three α values (1, 0.7 and 0.2) to vary the interaural coherence of the masker. The masker (SSN or VS) was simulated at

[13] The amount of reverberation was chosen to have a good sampling of binaural unmasking advantage. The anechoic room was considered to have the highest binaural unmasking advantage because noise interaural coherence is equal to one. The other levels of reverberation were chosen to significantly vary the noise interaural coherence and thus varying the expected binaural unmasking advantage. The room used by Lavandier and Culling (2010) satisfied all these requirements: they simulated rooms that led to a (broadband) binaural unmasking advantage of 3.5, 2.5, and 1.5 dB. The material was already generated and available, which was convenient to save time.

the same position across conditions (at 6.16 m and +16.4° relative to the listener) and played at a constant level of 60 dB SPL. The sentences were always anechoic and simulated at 2 m from the listener at azimuths of -90°, +20° or +40°. Their level was varied using an adaptive procedure (see below). To investigate the effect of audibility, two additional noise levels of 40 and 50 dB SPL were used for the single condition where the target was placed at -90° and α was equal to 1. This spatial configuration was chosen because it was expected to maximize the binaural unmasking advantage, since the noise coherence and the difference in azimuth between target and noise were maximized. Both noises, SSN and VS were used at those two levels.

Individual, linear amplification following the NAL-RP prescription formula (Dillon, 2012, chapter 10) was applied to the stimuli played to the HI listeners to compensate partly for their hearing loss.

V.2.3.2 Procedure

The experiment took place in a sound-attenuating test booth. The listener was seated on a sofa-chair with headphones on, and the experimenter was in front of a computer screen. Two lists of 16 sentences were used for each SRT measurement. The software played one sentence at a time along with the masking noise, which was played continuously at a fixed level. The listener was asked to repeat what he/she understood after each sentence and the experimenter marked on the software interface the number of correct words. The SNR for the first sentence was set to 4 dB. Then, the sentence level was varied according to the number of correct words for the previous sentence to converge towards the level at which the listener understood half of the words in the sentence. The speech level was increased if the number of correct words was below 50% and decreased otherwise. For the first five sentences, the target level was varied by 4-dB steps (initial phase). The step size decreased to 2 dB when at least 5 sentences had been presented scored and a reversal towards higher SNR had been observed, which indicated the start of the measurement phase. The measurement ended when at least 17 sentences were scored (i.e., at least 22 sentences were played considering the initial and measurement phases) and the standard error of the SNR across the measurement phase was below 1 dB, or when all 32 available sentences had been presented. The measured SRT was obtained by averaging the SNRs of the measurement phase. To avoid order and speech material effects, the conditions were presented in a random order and the 2 sentence lists were chosen randomly for each condition. No list was not played more than once to each listener.

A short break was given after each condition and the listener was allowed to have a longer break at any time in between conditions. If the experiment was done in one appointment, the participant was asked to have at least a 15-min break half way (i.e., after the 11th condition).

Table V.a summarizes the conditions tested. The measured SRTs were split into two subsets for the statistical analysis. Subset I contained the SRTs measured when varying the target azimuth and the absorption coefficient of the room for the masker (3 target azimuths x 2 masker types x 3 levels of reverberation = 18 conditions). Subset II contained the SRTs measured at the three levels (3 levels x 2 masker types = 6 conditions).

Subset	Masker type	Masker presentation level (dB SPL)	Target azimuth	Room absorption coefficient for the masker (α)
I	SSN, VS	60	-90°, 20°, 40°	1, 0.7, 0.2
II	SSN, VS	40, 50, 60	-90°	1

Table V.a: Summary of the experimental design: each column represents a tested factor. Subset I contains the conditions investigating the effect of the room and the target azimuth on binaural unmasking. Subset II contains the conditions assessing the effect of level on binaural unmasking. The two SRTs measured when the target was placed at -90° with the anechoic noises (SSN or VS) played at 60 dB SPL were common to the two subsets.

V.2.4 Model

V.2.4.1 Original model: Vicente et al. (2020).

The model considered in the present study and named here "Vic20" was proposed by Vicente et al. (2020, and presented in Chapter IV). It was shown to predict binaural speech intelligibility in the presence of stationary and envelope-modulated noises for NH and HI listeners. The target and masker signals at the listener's ears are taken as inputs to the model. They are calibrated to the sound level (in dB SPL) used during the experimental conditions (here, the sound level of the masker that was fixed during the adaptive measurement). To take into account hearing loss, the pure-tone audiogram of each listener is used as model input to spectrally shape an internal noise spectrum separately at each ear.

Based on the incoming signals and the internal noise spectra, the SNR at the better ear and the binaural unmasking advantage are computed for each time frame and frequency band combining the target's long-term characteristics with the masker's short-term characteristics (i.e., magnitude spectra at each ear, ITDs, and interaural coherence). The noise signal is windowed using 24-ms half-overlapping Hann windows to predict the better-ear SNRs and taking into account monaural dip-listening (Vicente and Lavandier, 2020). To predict binaural unmasking advantages, 300-ms long, half-overlapping Hann windows are applied to the signals to take into account binaural sluggishness (Hauth and Brand, 2018; Vicente and Lavandier, 2020). The target and windowed noise signals are passed through a gammatone filterbank with two filters per ERB ranging from 30 Hz to 19.9 kHz.

To compute the SNR at the better ear, the SNR at each ear is determined by the higher level between the external and internal noise and limited to 20 dB. The higher SNR across ears is selected as the better-ear SNR. The binaural unmasking advantage is computed only if the external signal levels are higher than the internal noise levels at both ears. It is obtained by applying the binaural masking level difference formula of Culling et al. (2004, 2005, Eq. V.1). The resulting better-ear SNRs and binaural unmasking advantages are averaged across time, integrated across frequency using a SII weighting (derived from the band importance function in Table 1 of ANSI S3.5, 1997), and added to obtain a binaural "internal" ratio. Differences between binaural ratios can be directly compared to differences between SRTs measured in listening tests.

In order to derive predicted SRTs, the binaural ratios are inverted (the higher the binaural ratio the better the intelligibility) and their offset is removed by first subtracting their mean and then adding the average SRT of the experiment,[14] which was chosen here as a reference to compare data and predictions. The predicted SRTs resulting from this transformation have the same mean value as the measured SRTs and the same relative differences as the initial inverted ratios.

The internal noise spectrum is defined by the formula proposed by Vicente et al. (2020), which, following the findings of Bernstein and Trahiotis (2008), combines an internal noise floor spectrum with a component that increases when the level of the external stimulus increases. The model computes the internal noise spectrum level in each frequency band based on the listener's audiogram for each ear (that is divided in two contributions interpreted as rough estimates of the OHC and IHC loss) and the external noise long-term broadband level (used here as a proxy of the level of the external stimuli). The formula contains three parameters, which were defined by Vicente et al. (2020) and were not modified for the current study.

V.2.4.2 Revised jitters: toward stimulus-dependent σ_ϵ and σ_δ

Vicente et al. (2020) showed that the effects of audibility and low sensation level on better-ear listening were well predicted by the "Vic20" model. In the present study, it was therefore assumed that any effect of low presentation level that is not accounted for by the model is due to a error in prediction of the binaural unmasking advantage. This was observed here for NH listeners, for whom the effect of low presentation levels on binaural unmasking was not well predicted (see Sec. V.3 and Fig. V-5). In order to address this discrepancy, a revision of the

[14] The SRTs are averaged across listeners within each group and then averaged across conditions and groups, giving the same weighting to each group regardless of their number of listeners. The same approach is used for the binaural ratios when removing their mean.

binaural unmasking component of the model is proposed that is based on a suggestion made by Durlach (1972). Even though in his original EC-model he described the jitters σ_ϵ and σ_δ as signal independent, he mentioned that they most likely would depend on the external noise level. Hence, a new characterization of the jitters is suggested here that is based on three assumptions: (1) The jitters increase when the level of the external stimuli decreases because the sounds are less audible and the auditory system makes more ILD and ITD alignment errors. (2) The level of the external stimuli can be approximated by the broadband level of the noise averaged across ears, which additionally makes the jitters frequency independent, as suggested Durlach (1963, 1972) and Wan et al. (2010). However, this level can still vary across listeners because of the listener-dependent NAL-RP amplification applied to the stimuli for the HI listeners. (3) The values of the jitters cannot be lower than their original values. Following these assumptions, the two jitters were re-defined as:

$$\sigma = max\left[1, 1 + a(N_{ref} - N)\right]\sigma_{Original} \tag{V.3}$$

where σ represents either the amplitude (σ_ϵ) or time (σ_δ) jitter, $\sigma_{Original}$ is the original value of σ (0.25 for σ_ϵ and 0.105 ms for σ_δ), N is the broadband level of the noise averaged across ears and a and N_{ref} are free parameters of the model. The predictions presented below are obtained with the parameter values leading to the best fit, namely, a = 10%/dB and N_{ref} = 50 dB.

V.2.4.3 Model evaluation

The target signal used for both models was the average of 128 randomly selected target stimuli from the experiment. The noise signal at the input to the model was a 2-min long sample of the noise played during the experiment.

The free parameters a and N_{ref} were varied within the range [1;10] %/dB and [45;60] dB, respectively, to fit the predictions to the data of the present study, especially for Subset II involving the different presentation levels (Table V.a and Fig. V-5). The prediction performance of the model was evaluated using the correlation r, $MeanErr$ and $MaxErr$ (see Sec. II.2 for computation details). When the free parameters were varied, the same values of a and N_{ref} were applied to both jitters to minimize the number of free parameters. The value of 10%/dB for a and of 50 dB for N_{ref} led to the best fit between data and predictions and were kept as final values. Applying these values, the model outputs did not change for the test conditions at 60 dB SPL (Subset I, Table V.a) between the original model and the model with the revised jitters. The model predictions obtained with the revised jitters only are displayed in the subsequent result figures (Figs. V-3-V-5), except for the figures related to the influence of presentation level (Fig. V-5), where the predictions of the original model are additionally shown for comparison purposes. A further comparison of the models and their ability to predict the effect of stimulus level is available in Appendix C.

V.3 Results

V.3.1 IPD sensitivity measurements

A Welch test performed between the two sessions collecting IPD_{freq} data (i.e., the upper frequency limit for detecting an IPD of 180°) revealed no significant difference for either NH ($p > 0.05$) or HI ($p > 0.05$) listeners. Hence, only the geometric-average IPD_{freq} across sessions was considered in the following analysis. Figure V-2 shows the resulting IPD_{freq}, which was equal to 1,257 Hz on average across NH listeners and equal to 665 Hz on average across HI listeners (individual data are shown as little black circles). A significant difference of IPD_{freq} between NH and HI listeners was confirmed using a Welch test ($p < 0.01$, statistical power = 0.99). The values of IPD_{freq} was not significantly correlated with listener's age within each group (statistical power for NH listeners was equal to 0.07 and to 0.31 for HI listeners) and no significant correlation was found between 4FAHL and IPD_{freq} for HI listeners ($p > 0.05$, statistical power equal to 0.05).

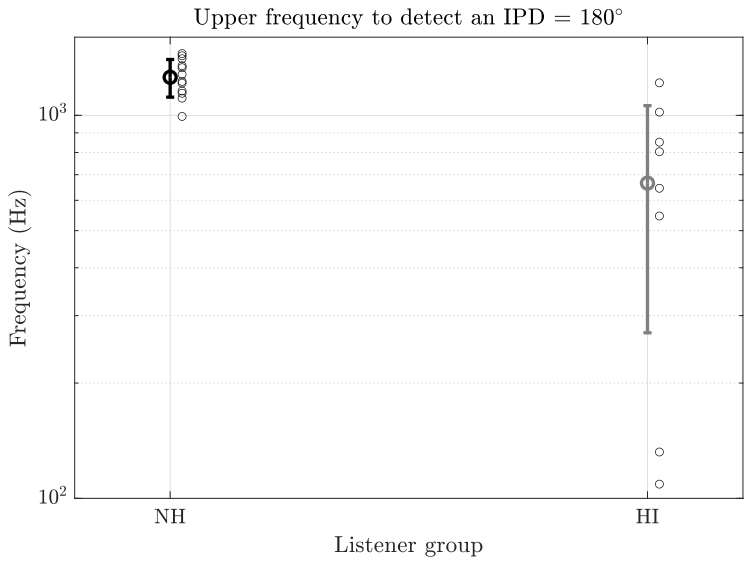


FIGURE V-2: Mean IPD_{freq} with ± 1 standard deviations across each listener group as a function of listener group. The little black circles show the individual data.

V.3.2 Effect of azimuth, reverberation, and masker type on SRTs

V.3.2.1 Measured and predicted SRTs

Figure V-3 shows the mean SRTs measured for the conditions at 60 dB SPL (Subset I, Table V.a). The SRTs for the HI listeners are higher on average than those for the NH listeners, as expected. The target was more intelligible in the presence of VS compared to a SSN. For a given level of reverberation applied to the masker (α), the SRT increased when the difference in azimuth between target and masker sources decreased, i.e., the SRT increased when the target was moved from -90° to +40° and from +40° to +20°. For the target placed at -90°, the anechoic maskers ($\alpha = 1$) led to lower SRTs than the reverberant maskers ($\alpha = 0.7$ or 0.2).

A linear mixed-effects model was designed, defining target azimuth, room absorption coefficient, masker type, listener group and all their interactions as fixed effects and a subject-specific intercept as random effect. This confirmed a significant effect of listener group [$F(2, 18) = 14.0$, $p < 0.01$], masker type [$F(1, 324) = 454.3$, $p < 0.0001$], target azimuth [$F(2, 324) = 184.0$, $p < 0.0001$] and reverberation level applied to the masker [$F(2, 324) = 31.3$, $p < 0.0001$]. The analysis also reported significant interactions that were further analyzed using Tukey pairwise comparisons (designed for multiple comparisons) to investigate the conditions that drove the significant effects. A summary is proposed in Table V.b.

The model predictions obtained with the revised jitters are plotted in Fig. V-3 as solid lines with the same color pattern as the data. The model is able to account for the influence of the target azimuth and the absorption of the room used for the masker. This is confirmed by good performance statistics ($r = 0.92$; $MeanErr = 0.9$ dB; $MaxErr = 2.0$ dB). However, the model slightly overestimates the difference between the SSN and VS maskers.

V.3.2.2 Measured and predicted binaural unmasking advantages

The target and masker sources were almost aligned when the target was placed at +20° (the masker was always at +16.4°). Since in this condition binaural unmasking was minimal, it was used in the following as the reference condition to evaluate the binaural unmasking advantage achieved by the different listeners. Note that the variations in better-ear listening across conditions were limited by the absence of the head in the simulation, so that the differences in SRT mostly reflect the differences in binaural unmasking advantage. Figure V-4 shows the binaural

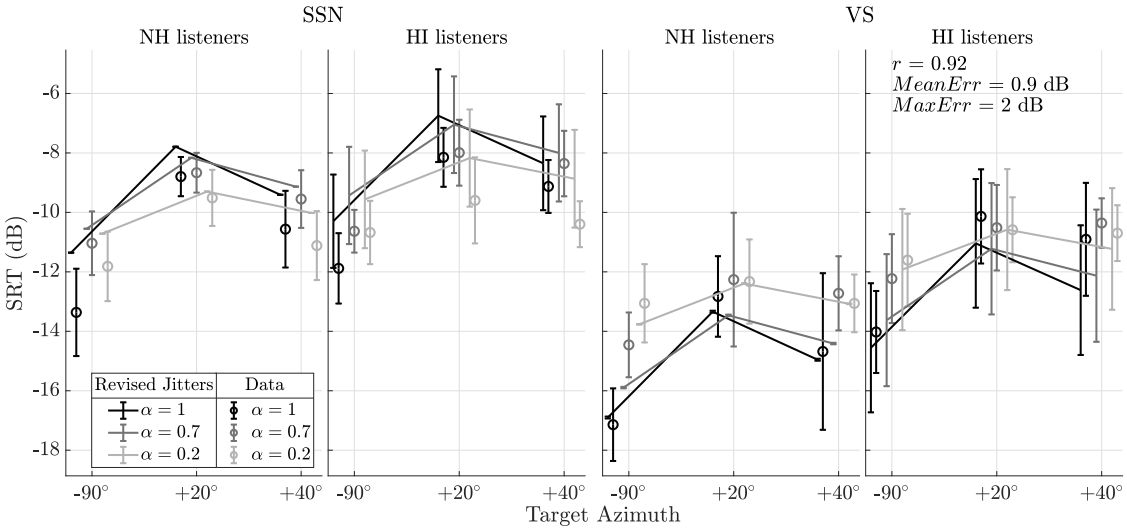


Figure V-3: Mean measured (circles) and predicted (solid lines) SRTs with ± 1 standard deviation across listeners as a function of target azimuth. Each panel contains data and predictions for a listener group and a masker type. The SRTs collected with the different α values (1, 0.7, 0.2) are plotted as dark, medium and bright greys. The noise was simulated at 6.16 m, +16.4° from the listener in a virtual room with an absorption coefficient α and played at 60 dB SPL. The target was anechoic and at 2 m from the listener.

unmasking advantages computed as the SRT at +20° minus the SRTs at the other target azimuths, as a function of target azimuth. The binaural unmasking advantage increased when the target was moved from +40° to -90°, i.e., when the difference in source azimuth increased. The binaural unmasking advantage at +90° decreased when reverberation increased.

A similar linear mixed-effects model as above was applied to analyze the observed binaural unmasking advantage. There were significant effects of listener group $[F(1,18) = 6.7,\ p < 0.05]$, target azimuth $[F(1,214) = 88.5,\ p < 0.0001]$, room absorption coefficient $[F(2,214) = 36.3,\ p < 0.0001]$ and masker type $[F(1,214) = 7.1,\ p < 0.01]$. The significant interaction and the related pairwise comparisons reported by the linear mixed-effects model are summarized in Table V.c.

The binaural unmasking advantages predicted by the model with revised jitters are shown in Fig. V-4. As for the SRTs, the influences of target azimuth and reverberation on binaural unmasking are accurately predicted ($r = 0.92$; $MeanErr = 0.5$ dB; $MaxErr = 1.1$ dB). When the target is placed at -90°, the model predicts the same binaural unmasking advantage, for both listener groups and masker types, which is in line with the data. Note that the model predicts very similar binaural unmasking advantages for all listeners within each group, which results in very small standard deviations compared to those for the data.

The correlation between the results of the TFS-AF test (IPD_{freq}) and the mean binaural unmasking advantage across conditions was significant for the two groups combined ($r = 0.50, p = 0.02$) but not for each group separately. The correlation between IPD_{freq} and the SRTs in the condition that maximized binaural unmasking advantage (SSN, target azimuth = -90°, $\alpha = 1$) was not significant either within each group or for the two groups combined.

V.3.3 Effect of presentation level on SRTs

Figure V-5 displays the SRTs measured in the anechoic conditions with maximal masker-target separation as a function of masker level (Table V.a, Subset II). The SRT decreased when the masker level increased and when the target was masked by a VS compared to a SSN. The HI listeners had higher SRTs than the NH listeners.

Listener group x Masker type		
	$[F(1,324) = 44.7,\ p < 0.0001]$	
HI, NH	SSN > VS	$10^{-4},\ 10^{-4}$
VS	HI > NH	10^{-4}
Listener group x Room absorption coefficient		
	$[F(2,324) = 7.1,\ p < 0.001]$	
HI	0.7 > 1, 0.2	$10^{-2},\ 5 * 10^{-2}$
NH	0.7, 0.2 > 1	$10^{-4},\ 10^{-4}$
1, 0.7	HI > NH	$10^{-3},\ 5 * 10^{-2}$
Listener group x Target Azimuth		
	$[F(2,324) = 3.7,\ p < 0.05]$	
HI	+40°, +20° > -90°	$10^{-4},\ 10^{-4}$
NH	+20° > +40° > -90°	$10^{-4},\ 10^{-4}$
-90°, +40°, +20°	HI > NH	$10^{-2},\ 10^{-3},\ 5 * 10^{-2}$
Target azimuth x Room absorption coefficient		
	$[F(4,324) = 23.6,\ p < 0.0001]$	
-90°	0.2, 0.7 > 1	$10^{-4},\ 10^{-4}$
+40°	0.7 > 1, 0.2	$10^{-4},\ 10^{-4}$
1, 0.7, 0.2	+20° > -90°	$10^{-4},\ 10^{-4},\ 10^{-4}$
1, 0.7	+40° > -90°	$10^{-4},\ 10^{-4}$
Room absorption coefficient x Masker type		
	$[F(2,324) = 22.6,\ p < 0.0001]$	
1, 0.7, 0.2	SSN > VS	$10^{-4},\ 10^{-4},\ 10^{-4}$
SSN	0.7 > 1, 0.2	$10^{-4},\ 10^{-4}$
VS	0.2, 0.7 > 1	$10^{-4},\ 10^{-4}$

TABLE V.B: Summary of the significant interactions and their related Tukey pairwise comparisons for the subset of SRTs measured at 60 dB SPL (see Subset I in Table V.a). The first column set a condition, the second a contrast and the third gives the significance level. For instance, the first comparison indicates that for both HI and NH listeners, the SRTs are significantly higher when the target was masked by a SSN compared to a VS ($p < 0.0001$, for each group).

A linear mixed-effects model was designed to analyze the SRTs with presentation level (Subset II,Table V.a), listener group, masker type and all their interactions as fixed effects and a subject-specific intercept as random effect. There were significant effects of masker type $[F(1,93) = 129.2,\ p < 0.0001]$, masker level $[F(2,93) = 167.1,\ p < 0.0001]$ and listener group $[F(1,18) = 129.2,\ p < 0.01]$. The significant interactions and the related pairwise comparisons reported by the linear-mixed effect model are summarized in Table V.d.

The model predictions obtained with the revised jitters are displayed in Fig. V-5 as solid lines with the same color pattern as the data. The predictions of the model are accurate for both group of listeners ($r = 0.93$; $MeanErr = 1.1$ dB; $MaxErr = 2.0$ dB). Figure V-5 also shows (dashed lines) the SRTs predicted with the Vic20 model ($r = 0.85$; $MeanErr = 1.1$ dB; $MaxErr = 3.1$ dB) to show the benefit of the revised jitters, which better account for the effect of audibility, especially for the NH listeners. The predictions for the NH listeners at 40 dB SPL are 2 dB higher than for the Vic20 model. The predictions obtained with the HI listeners are slightly shifted up at 40 dB SPL but are still accurate. It is worth noting that the implementation of the binaural unmasking component of the Vic20 model was sufficient to predict the effect of presentation level on binaural unmasking for HI listeners. The predicted SRTs with the revised jitters are shifted a bit down (by 0.14 dB) at 60 dB SPL because of the offset that is applied to convert the binaural ratios into predicted SRTs.[15]

[15]The offset is based on the average binaural ratio across conditions, which changes with the modified jitters in the binaural unmasking component of the model because the binaural ratios at 40 and 50 dB SPL are influenced by this modification. While binaural ratios are not affected by the modification of the jitters at 60 dB SPL, the corresponding predicted SRTs are slightly offset due to the change in average binaural ratio.

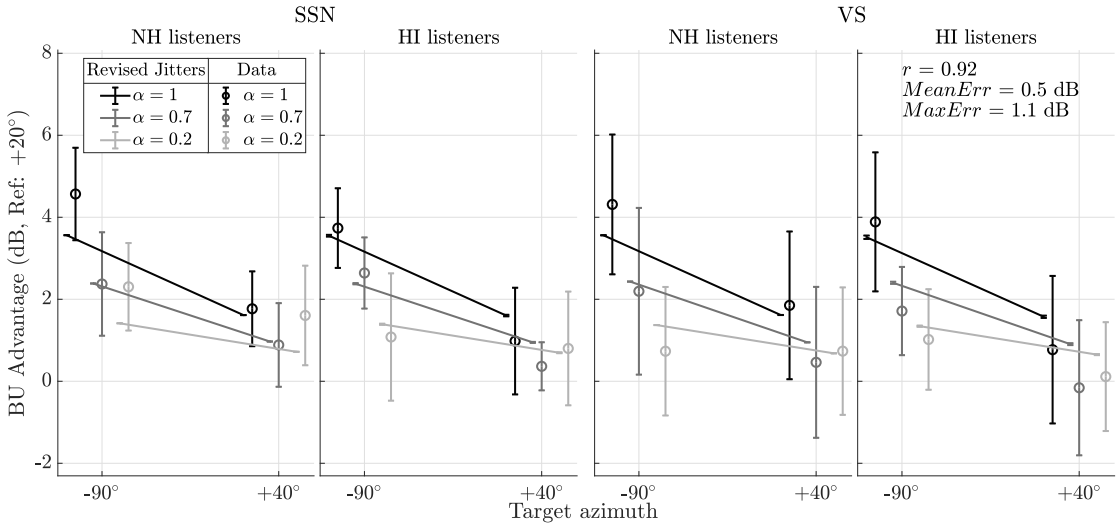


Figure V-4: Same as Fig. V-3 but for the binaural unmasking advantages obtained by subtracting the SRTs at -90° and +40° from the SRTs at +20°.

	Target Azimuth x Room absorption coefficient	
	$[F(2,214) = 14.3,\ p < 0.0001]$	
-90°	$1 > 0.7 > 0.2$	$10^{-4},\ 5 * 10^{-2}$
1, 0.7	-90° > +40°	$10^{-4},\ 10^{-4}$

Table V.c: Same as Table V.b but for binaural unmasking advantages measured at 60 dB SPL.

V.4 Discussion

V.4.1 Influence of hearing loss

Regarding the results of the TFS-AF test, the HI listeners had significantly lower IPD_{freq} values than the NH listeners (average values: 665 Hz vs 1,257 Hz), which is in line with the literature (e.g., Füllgrabe and Moore, 2017; Neher et al., 2011; Santurette and Dau, 2012). There was also a significant correlation between binaural unmasking advantage and IPD_{freq} for the two groups combined, but not for each group separately. This is in line with Santurette and Dau (2012), who also did not find a significant correlation between IPD_{freq} and binaural unmasking advantage for HI listeners. However, Strelcyk and Dau (2009), Neher et al. (2011) and Neher et al. (2012) found a significant correlation between IPD threshold and the SRTs measured with HI listeners in the conditions that maximized the difference in ITD between target and masker. This is inconsistent with the present study, where there was no significant correlation ($p < 0.05$, statistical power equal to 0.05) between IPD_{freq} and the SRTs measured for the HI listeners with the anechoic SSN (i.e., the noise maximizing binaural unmasking advantage) and the target at -90° (i.e., largest source separation). This difference might be explained by the fact that Strelcyk and Dau (2009) found a significant correlation only when they measured the IPD threshold in noise (while the present measure was in quiet). Neher et al. (2011) and Neher et al. (2012) measured their SRTs with speech maskers and tested listeners with more severe hearing loss (i.e., their listeners had an average 4FAHL that was 10 and 13 dB higher than in the present study).

Füllgrabe and Moore (2018) analyzed the results of 19 studies to investigate the relationship between age, hearing threshold and IPD sensitivity. They found that IPD threshold was negatively correlated with age and hearing threshold but more so for age. They also found an interaction between age and hearing threshold: for listeners older than 58 years, the effect of age

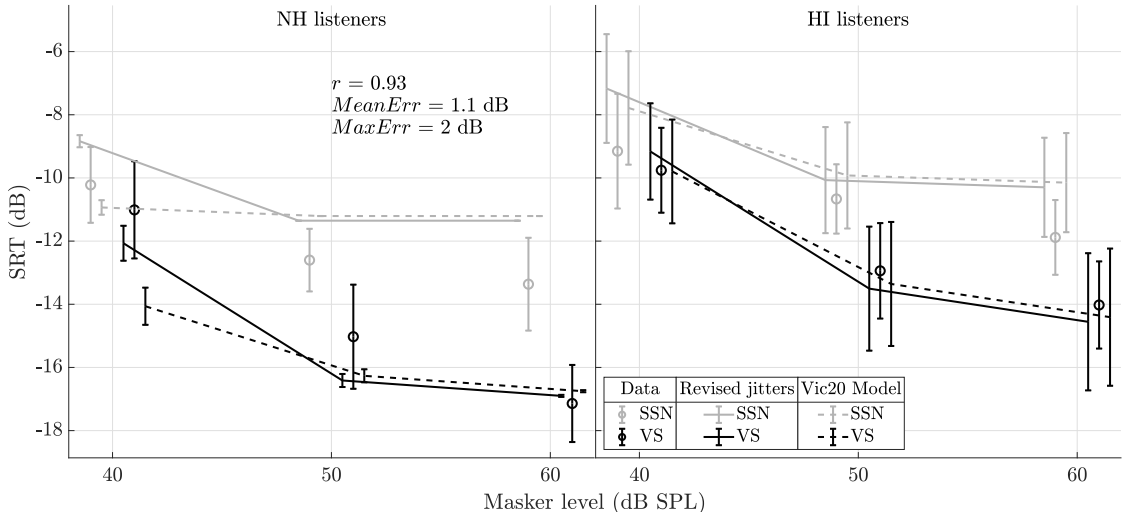


FIGURE V-5: Mean measured (circles) and predicted (lines) SRTs with ± 1 standard deviation across listeners as a function of masker level. The measured and predicted SRTs obtained with the SSN and VS maskers are plotted in grey and black, respectively. Each panel is for one listener group. The target was placed at 2 m, -90° from the listener while the (anechoic) masker was at 6.16 m, +16.4°. The SRTs predicted with the model with revised jitters are displayed as solid lines, while the dashed lines show the predictions of the Vic20 model. Performance statistics are displayed only for the model with revised jitters.

was greater for those who had a five frequency-average (0.25, 0.5, 1, 2, 4 kHz) hearing loss higher than 30 dB HL. In the present study, HI listener's age was not correlated with IPD_{freq}, which is in agreement with Santurette and Dau (2012) and Strelcyk and Dau (2009) but not with the results of Neher et al. (2011) and Neher et al. (2012) or with the findings of Füllgrabe and Moore (2018). This discrepancy might be driven again by the differences in hearing threshold between studies, given that the mean age was higher than or equal to 58 years in all studies.

The binaural unmasking advantages measured here for the NH listeners were only slightly but significantly higher (by 0.6 dB, $p < 0.05$) than those for the HI listeners. This difference is in general agreement with George et al. (2012), who measured the binaural unmasking advantage[16] for NH and HI listeners using an anechoic SSN, an anechoic modulated SSN and two diotically-reverberated modulated SSNs (in these conditions reverberation was also applied to the speech). They found that HI listeners had significantly lower binaural unmasking advantage of about 2-4 dB versus 5-6 dB for NH listeners. The difference between HI and NH listeners was larger than observed in the current study, which may again be explained by the more severe hearing losses of their listeners (i.e., their average 4FAHL was 45 dB HL).

Goverts and Houtgast (2010) measured binaural unmasking advantages[16] for 6 NH and 25 HI listeners. The NH listeners had mean binaural unmasking advantage of 5.3 dB (standard deviation of 0.8 dB) and 17 of the HI listeners had binaural unmasking advantage within ±2 standard deviations, indicating no significant difference in binaural unmasking advantage between the NH and these HI listeners. For the eight remaining HI listeners, the binaural unmasking advantages were smaller, which the authors mainly attributed to a loss of phase and time coding accuracy. Hence, they suggested that a lack of temporal coding related to hearing loss (so-called "supra-threshold deficits") can affect binaural unmasking for speech for some HI listeners but not for all. In the present study, the IPD_{freq} measure—that can be seen as a measure of a lack of temporal coding—was correlated with the mean binaural unmasking advantage for the two groups combined, but not within each group. This means that the IPD_{freq} measure was appropriate

[16]The binaural unmasking advantage was defined as the difference in SRT between diotic listening and a dichotic listening condition where the speech was presented out-of-phase and the noise in-phase at the ears.

Masker level x Masker type		
	$[F(2,93) = 15.3,\ p < 0.0001]$	
50, 60	SSN > VS	$10^{-4},\ 10^{-4}$
SSN	40 > 50 > 60	$10^{-4},\ 5 * 10^{-2}$
VS	40 > 50 > 60	$10^{-4},\ 10^{-4}$
Masker level x Listener group		
	$[F(2,93) = 129.2,\ p < 0.05]$	
50, 60	HI > NH	$5 * 10^{-2},\ 10^{-2}$
HI	40 > 50 > 60	$10^{-4},\ 5 * 10^{-2}$
NH	40 > 50 > 60	$10^{-4},\ 10^{-4}$

Table V.d: Same as Table V.b but for SRTs measured at 40, 50 and 60 dB SPL (see Subset II in Table V.a).

to partly explain the difference between the two groups but not sensitive enough to predict the individual differences amongst HI listeners (who were a rather cohesive group).

Bronkhorst and Plomp (1989) tested the ability of HI listeners with symmetric and asymmetric hearing loss to benefit from ILD and ITD to improve speech intelligibility in stationary noise under anechoic conditions. They presented speech target in front of the listener and a SSN at 0°, +30° or +90°. They showed that HI listeners with a symmetric hearing loss benefited from ITD as much as NH listeners when the masker was placed at +90°, but this was not the case for the asymmetric HI listeners (as further confirmed by Bronkhorst and Plomp, 1990). However, the benefit for the symmetric HI listeners was lower (by 1.2 dB) than for the NH listeners for the noise at +30°. This indicates an interaction between the effects of listener group and masker azimuth that was not highlighted in the present study (Table V.d).

Based on the literature described above, some conclusions on the relation between binaural unmasking and hearing loss can be drawn. The studies agree that a significant number of HI listeners show similar binaural unmasking to NH listeners, while some HI listeners show lower binaural unmasking. When a HI listener shows reduced binaural unmasking advantage, it is either due to a supra-threshold deficit or an asymmetric hearing loss. The statistical analysis of the SRTs measured here at 60 dB SPL showed a significant interaction between listener group and masker type. No difference between groups was found for SRTs measured with a SSN, while SRTs measured with the VS maskers were significantly higher for the HI listeners. This may be interpreted as a reduction of the ability to listen in the masker's gaps (Festen and Plomp, 1990). It is also in line with Jensen and Bernstein (2019), who suggested that the dip-listening advantage is reduced for HI listeners when audibility is not equalized across listener groups, which was the case here because the NAL-RP amplification only partly compensates for the reduced audibility due to hearing loss (Glyde et al., 2015).

V.4.2 Factors influencing the SRT

The highest SRTs occurred when the intermediate reverberation ($\alpha = 0.7$) was used with the SSN. This rather surprising observation may be explained by two interacting effects when adding reverberation to a noise signal (Lavandier and Culling, 2010). On the one hand, the interaural coherence of the noise decreases when reverberation increases, which leads to lower intelligibility. On the other hand, reverberation distorts the masker spectrum, thus creating spectral differences between the anechoic target and the reverberated masker, which can led to better intelligibility. Hence, apply reverberation to the masker when it was anechoic impaired intelligibility (i.e., the effect of coherence dominated from $\alpha = 1$ to $\alpha = 0.7$), while adding further reverberation improved intelligibility (i.e., effect of room coloration dominated from $\alpha = 0.7$ to $\alpha = 0.2$). These trends are confirmed by the model: decreasing α to 0.7, the better-ear SNR increases from 4.2 dB to 4.5 dB while the binaural unmasking advantage decreases from 1.8 to 1.2 dB. Further decreasing α to 0.2, the better-ear SNR increases from 4.5 to 5.6 dB and the binaural unmasking advantage decreases from 1.2 to 0.8 dB.

The significant effect of masker type is demonstrated in Fig. V-3 by lower SRTs with the VS masker. The model slightly overestimates this difference, which was also observed by Vicente

et al. (2020). This is likely a misprediction from the better-ear listening component of the model because the overestimation is observed even when the sources are almost aligned (or co-located in Rana and Buchholz, 2016), i.e. when there is basically no binaural unmasking advantage. This explanation is supported by the model predictions of the benefit of the VS modulation at the target azimuth of +20°, which in the measured data was 2.7 dB while the model predicts 4.1 dB of improvement, 4.0 dB of which are linked to the better-ear listening component.

For the anechoic measurements, SRTs increased when the sensation level decreased. This could be due to a decrease of target audibility, a decrease of binaural unmasking advantage and/or, for the VS masker, a decrease of dip listening advantage. Given that the Vic20 model accurately predicts the effect of audibility on the better-ear SNR (Vicente et al., 2020), while its predictions are not accurate for the lower level, the decrease in the observed SRT with decreasing audibility is likely due to a decrease of binaural unmasking advantage.

V.4.3 Factors influencing binaural unmasking

The binaural unmasking advantage was lower for the VS masker than for the SSN (by 0.5 dB). During the dips of the VS masker, the binaural unmasking advantage is probably reduced because there is less masker energy to cancel. As a consequence, binaural unmasking advantage is reduced when averaged over time. A similar result was found by George et al. (2012), who reported a 1-dB reduction in binaural unmasking advantage[16] with a speech modulated SSN compared to SSN.

Reverberation on the masker also influenced the measured binaural unmasking advantage. This is because, when reverberation increased as the absorption coefficient of the room was reduced, the interaural coherence of the noise decreased (because reverberation is different at the two ears). As a result, the masker became more difficult to cancel (Durlach, 1972; Licklider, 1948) and the binaural unmasking advantage was reduced (Lavandier and Culling, 2010).

The target azimuth significantly affected the binaural unmasking advantage for both groups of listeners. The larger the absolute difference in target and masker azimuth, the better the intelligibility, and the larger the binaural unmasking advantage. The largest binaural unmasking advantage (anechoic SSN, target at +90°) of 4.6 dB is in good agreement with values reported in the literature (Lavandier and Culling, 2010; George et al., 2012; Goverts and Houtgast, 2010; Bronkhorst and Plomp, 1988, 1989).

The success of the revised model in predicting binaural unmasking at low presentation levels (in particular at 40 dB SPL) for NH listeners, which was not the case for the Vic20 model, suggests that binaural unmasking was influenced by the presentation level of the masker (as discussed in Sec. V.4.2). Not clear was not observed for HI listeners most likely due to the amplification that was provided to compensate for their hearing loss (which increased the broadband stimulus level) combined with their rather mild hearing impairment. Hence, to further understand the effect of sensation level on binaural unmasking, listeners with more severe hearing loss need to be tested. Additionally, listeners with asymmetric hearing loss should be investigated, as they are known to generally show lower binaural unmasking advantages (Bronkhorst and Plomp, 1989, 1990), likely due to different sensation levels at the two ears (Durlach et al., 1981).

V.4.4 Predicting binaural unmasking

The binaural unmasking advantage at 60 dB SPL is well predicted by both the Vic20 model and the revised model ($r = 0.92$, $MeanErr = 0.5$ dB, $MaxErr = 1.1$ dB for both models). For lower noise levels, the Vic20 model predicts the effect of level on SRTs for the HI listeners but not for the NH listeners, especially at a noise level of 40 dB SPL. This is most likely due to poor prediction of the binaural unmasking component. To redue this discrepancy, the present model accommodates the effect of audibility on binaural unmasking by increasing the ITD and ILD jitters at low presentation levels. Given that the jitters reflect the amplitude and time alignment errors in the equalization stage involved in the EC mechanism, their increase when stimulus audibility decreases seems plausible (Durlach, 1972). However, even though such a revision showed promising results here, it needs to be tested in more conditions, including with asymmetric HI listeners as well as listeners with more severe hearing loss.

Beutelmann and Brand (2006) and Beutelmann et al. (2010) developed a binaural model for predicting speech intelligibility in unmodulated and modulated noise for NH and HI listeners, which also includes a level-dependent EC mechanism. Their model includes an internal noise that is spectrally matched to the individual's audiogram and uncorrelated across ears, which is added to the external noise. This results in an interaural coherence of the combined (internal+external) noise that is lower than the external noise coherence, thus limiting the efficiency of the EC mechanism: the lower the external noise level, the lower the combined noise interaural coherence, and the lower the efficiency of the EC mechanism. Such an approach is in line with the data of the present study, as well as with Bernstein and Trahiotis (2017), who also assumed that the coherence of the combined (internal+external) noise is lower than the coherence of the external noise in a model that predicts tone detection thresholds in noise (as suggested by Bernstein and Trahiotis, 2008). In their model, the predictions needed to be fitted to the data for each tone frequency and configuration of tone and noise by varying the parameter that defines the internal noise level, in order to optimize the combined (internal+external) noise interaural coherence. Interestingly, Bernstein and Trahiotis (2020) showed that the values of the fitting parameters are consistent across different studies.

The models of Beutelmann and Brand (2006), Beutelmann et al. (2010) and Bernstein and Trahiotis (2017) differ in two major ways from the current model. First, the level dependence of the EC mechanism is implemented here within the jitters involved in the equalization stage, while the above studies take the level dependence into account through the combined noise interaural coherence. Second, their models include a binaural internal noise while here the internal noise levels are defined independently at each ear without any *true* binaural component to them. These levels do not influence the interaural coherence of the external noise, but they reduce the binaural unmasking advantage as soon as the external level at one of the ears is below the corresponding internal noise level.

Beutelmann and Brand (2006) and Beutelmann et al. (2010) also implemented ITD and ILD jitters in their EC mechanisms that were dependent on frequency; assuming an EC mechanism that operates independently within each frequency channel. This is in contrast to the present implementation of the jitters, whose magnitudes are determined by the broadband level of the external stimuli. However, even though the present approach does not directly consider the stimulus level within frequency bands, it still takes into account the effect of audibility on binaural unmasking as a function of frequency because the binaural unmasking advantage is set to 0 if the external stimulus levels are below the frequency-dependent internal noise levels.

Beutelmann et al. (2010) suggested that the jitters may also be time dependent and vary with the short-term level of the noise signal. With respect to the present study, such an approach might help to explain the measured difference in binaural unmasking advantage between the SSN and VS maskers. When the VS masker short-term level decreases, the jitters would increase and thus, on average across time, the jitters would be higher for the VS masker than for the SSN, which would lead to lower binaural unmasking advantage. However, when implementing such time-dependent jitters in the present model, the 300-ms window used to take into account binaural sluggishness in the binaural unmasking stage removed most of the amplitude modulations of the VS masker. As a consequence, the predicted difference in binaural unmasking advantage between VS and SSN was negligible (0.02 dB).

The work of vom Hömel (1984, cited in Beutelmann and Brand, 2006) suggests that the jitters involved in the EC mechanism depend on the ILD and ITD of the external noise signal, such that the equalization stage is less accurate when the ILD and ITD increase. Similarly, Bernstein and Trahiotis (2018) further developed the detection model of Bernstein and Trahiotis (2017) by adding a jitter that reflects the auditory system's ability to equalize an ITD. This jitter increased with increasing ITD and the rate of growth varied for each tested frequency. The revised jitters proposed here depend only on the long-term, broadband level of the stimuli and are not affected by the ILD and ITD. This was sufficient to predict the present data, and for the sake of simplicity, no more modifications were added to the model.

V.5 Summary

This experimental and modelling study investigated the effect of hearing loss on the contribution of binaural unmasking to speech intelligibility in noise. The effects of four variable were tested:

the difference in azimuth between masker and target, audibility via hearing loss and presentation level, masker type, and reverberation on the masker. They all had a significant effect on intelligibility, with HI listeners showing overall slightly reduced binaural unmasking advantage compared to the NH listeners. The model proposed by Vicente et al. (2020, and presented in Chapter IV) accurately predicted all the measured effects on both SRTs and binaural unmasking advantages at a moderate stimulus level (60 dB SPL), but failed to predict the effect of lower presentation levels. To address this limitation, the jitters used in the binaural unmasking component of the model needed to be revised. This revised implementation of the model needs to be further tested on conditions not used to define its parameters as well as with HI listeners with a broader range of hearing losses, particularly, listeners with more severe hearing loss at low frequencies and asymmetric hearing loss.

VI Concluding remarks and outlook

VI.1 Summary

Throughout the three studies, the binaural speech intelligibility models proposed by Collin and Lavandier (2013) and Lavandier et al. (2018) were further tested and developed towards a single binaural model that predicts speech intelligibility in envelope-modulated noises for NH and HI listeners. Overall, 154 acoustic conditions (100 in Chapter III, 32 in Chapter IV and 22 in Chapter V) split into 12 experiments were modelled during the project (including the datasets used to verify the backward compatibility presented in Chapter IV). The model accuracy across datasets is consistently high with r ranging from 0.85 to 0.98, $MeanErr$ from 0.4 to 1.4 dB and $MaxErr$ from 0.6 dB to 7.1 dB.

The model from Collin and Lavandier (2013) was tested in Chapter III. A method inspired by a sensitivity analysis (adapted from Saltelli et al., 2010) was used in order to quantify the influence of the parameters involved in the model and set the parameter values that led to the best fit between data and predictions across experiments. Chapter III also considered the influence of binaural sluggishness in the binaural unmasking component of the model by increasing the duration of the temporal window that frames the masker signal to predict the binaural unmasking advantage. The study also showed that better-ear glimpsing most likely requires short-time temporal resolution to benefit from the masker modulations, however, it is somehow limited by the binaural sluggishness of the auditory system. Implementing these two temporal resolutions within the better-ear listening component of the model would need further consideration.

Chapter IV proposed a modification of the model developed by Lavandier et al. (2018), which allowed to predict intelligibility for NH and HI listeners at various sensation levels using a single model version. This was accomplished by implementing an internal noise level at each listener's ear to simulate hearing impairment. Some binaural tone detection models (e.g., Bernstein and Trahiotis, 2018) as well as binaural speech intelligibility models (e.g., Beutelmann and Brand, 2006) already consider such an internal noise, however, the implementation proposed in the present study is new in the literature. It considers a distribution of the audiogram in OHC and IHC loss to create an internal noise floor (based on the OHC loss) as well as a stimulus-dependent internal noise level (based on the IHC loss) that increases when the external stimulus level increases. The concept of considering the internal noise in two components was already suggested by, e.g., Bernstein and Trahiotis (2008) but the current study has substantially extended the scope of application of this definition. This means that hearing impairment can be considered as a noise relying on the external sound level in the auditory pathway, limiting abilities to analyze and segregate speech from noise.

Chapter V investigated the contribution of binaural unmasking to speech intelligibility for NH and HI listeners. The experimental data collected especially for the study demonstrated that HI listeners showed lower advantage due to binaural unmasking compared to NH listeners at 60 dB SPL. Also, the decrease of binaural unmasking advantage when sensation level decreases was confirmed by the collected SRTs. The outcomes of the modelling section were also valuable. First, the formula developed by Durlach (1972, which was further utilized by Culling et al., 2004, 2005) was modified to consider the effect of presentation level. To do so, the fixed interaural jitters were replaced by jitters that varied as a function of the level of the external stimuli. This highlights the fact that the auditory system would make more errors in interaural time and level alignments when sensation level decreases. Second, applying the modified formula of Durlach (1972) only when the signals are audible within a frequency band (i.e., when the external signal levels are above the internal noise level at each ear) allowed to predict binaural unmasking

for NH and HI listeners. Third, considering an internal noise without binaural component (as opposed to, e.g., Bernstein and Trahiotis, 2017 or Beutelmann and Brand, 2006) was sufficient to predict, at least, the present data on binaural unmasking.

VI.2 Model strengths and weaknesses

The knowledge on the auditory system has been extended through the project thanks to modelling and experimental studies. Especially, the studies have allowed to better understand the mechanisms of better-ear listening and binaural unmasking and their relationship with hearing loss. The final model version (developed in Chapter V) is able to predict the advantage due to better-ear listening and binaural unmasking for NH and HI listeners. This is mainly realized by modelling the listener's hearing loss by an internal noise level based on the audiogram. It is defined independently at each ear (i.e., without considering any binaural processing) and composed of two components, namely, an internal noise floor and a stimulus-dependent internal noise. This is a major contribution of the model to the field, because other competing models (e.g., Beutelmann et al., 2010; Schädler et al., 2018) do solely consider an internal noise floor. Moreover, it was also shown that the internal noise floor is most likely linked to OHC loss while the stimulus-dependent internal noise is linked to IHC loss. This suggests that any future behavioural model that aims to emulate the auditory signal processing of speech in noise should consider these two different types of internal noises along the auditory pathway. Finally, hearing loss was modelled only using the listener's audiograms and the broadband external stimulus level. This means that other deficits such as reduced ITD sensitivity, reduced temporal resolution, or reduced frequency resolution were not required as an input to the model, which simplifies its implementation and application. However, given the variability seen in the individual data, it may well be that some of the differences seen between subjects are explained by an impairment in these additional processes and therefore, they may need to be considered in a future model version.

Another strength of the model is in its approach to predict speech intelligibility. The model assumes that the auditory system processes ILD and ITD independently. This approach differs from the other competing models, especially the one from Beutelmann et al. (2010) that considers ILD and ITD processing dependent on each other. It is important to have different model concepts to better understand the underlying auditory processes. Using its own concept, the current model is accurate at predicting the diverse datasets considered during the project, since the mean absolute errors between data and predictions are always about 1 dB and the Pearson's correlation coefficients are always higher than 0.9. Furthermore, when plotting individual measured SRTs versus individual predicted SRTs (see Chapter IV), the model was accurate in presenting the general trends. However, since the listeners in the different experiments were rather cohesive groups and the model parameters were defined using these experiments, it can not be concluded that the model is able to successfully predict speech intelligibility in individual listeners. Nevertheless, it is an encouraging start for a next model version that aims at predicting individualized speech intelligibility.

The model can be criticized also on different aspects. First, it was tested at predicting SRTs only. This is not representative of daily life in which people understand not only 50% of target words. The model would need to be tested at predicting psychometric functions giving the relationship between SNRs and percentage of correct words. Furthermore, the model inputs are the clean speech and the noise alone, which is not representative of real-life listening where the target+noise mixture is only available. Hence, developing a model that can use only the target+noise mixture with our model concept would be an interesting future goal. For instance, Mi and Colburn (2016) and Hauth et al. (2020) proposed two models that consider a blind processing (i.e., processing using only the target+noise mixture) to segregate the target speech from the noise signal. This might reflect what happens in the auditory system: the auditory system segregates the two signals to be able to further analyse speech. Moreover, the current model does not consider any top-down processing, which is related to the use of knowledge acquired by experience such as vocabulary. Josupeit and Hohmann (2017) and Schädler et al. (2018) developed models that compare templates of clean speech (that are generated prior predictions) to templates of noisy speech to predict intelligibility. The templates of clean speech can be seen as way to model the vocabulary that is stored somewhere in the brain. Such an implementation has not been required in the current model but it might be of interest in a future

version. Finally, the model is likely not able to predict some acoustic conditions or scenarios closer to real life (see Sec. VI.3), which would have to be highlighted in future works to justify model updates.

Some assumptions that have been made during the model development will have to be reconsidered. The assumptions in Chapter IV simplifying the internal noise implementation might need further consideration. For instance, the parameter η that defines the proportion of OHC loss is frequency-independent. It was an intentional simplification to reduce the number of fitting parameters of the model, which was sufficient to predict intelligibility in the considered datasets. However, it is likely not representative of the individual listener's hearing loss. It was also assumed that the level of the internal noise was defined using the external stimulus level, which was approximated by the level known during the experiment, i.e., the noise level. Then, this approximation was accurate only for negative SRTs. One way to avoid this assumption is to predict percentage of correct words so that the target and masker levels are known.

VI.3 Outlook

First, the model could be tested at predicting positive SNRs (similarly to Hauth et al., 2020), this would challenge the assumption that uses the broadband level of the external noise as a proxy of the external level. Moreover, according to the results of Smeds et al. (2015), SNRs encountered in important daily situations are often positive unlike the measured SRTs involved in the present studies that were negative. Another way to challenge this assumption is to test the model on signals that do not have the same spectra as the ones considered during the project, i.e., speech-like spectra. The broadband level of speech-like spectra is largely defined by the low-frequency energy while, for instance, the broadband level of a white noise is defined by all frequencies equally due to its flat spectrum. One could imagine that the proposed internal noise formula is only accurate for the same kind of noise with a speech-like spectrum. Hence, testing the model at predicting intelligibility of speech in different kind of noises could be interesting to further test the internal noise implementation.

The proposed model must be tested at predicting datasets collected with more severe and/or asymmetric HI listeners. This will allow to generalize its relevance to predict intelligibility for HI listeners. It is worth noting that the model is able to predict the average difference between groups of NH and HI listeners but it is yet to be confirmed that it can predict differences in individual intelligibility thresholds (even if the correlation analysis in Chapter IV is encouraging). However, the individual intelligibility performances are primarily driven by hearing loss and secondarily by cognitive and/or aging effect (as suggested by Akeroyd, 2008, and Glyde et al., 2011). This means that the model might need to consider these secondary effects to predict individual speech intelligibility.

The model needs to be tested at predicting conditions involving non-linear hearing-aid processes such as frequency lowering, noise suppression or non-linear amplification. The conditions considered by Kates and Arehart (2014) to develop the HASPI can be a source of inspiration. This would be an important step towards providing a model that can predict effects of hearing impairment and hearing aids on binaural speech intelligibility.

The model can be challenged at predicting datasets in which the target is reverberated. The model for NH and HI was tested only with conditions involving an anechoic target. This was done to avoid any effect of target smearing, however, this is not representative of real-life scenario. Leclère et al. (2015) modified the model of Lavandier and Culling (2010) to be able to predict the effect of target smearing. This modification is not implemented in the current model; hence, it is very likely that it cannot predict this effect. It is also worth considering dynamic scenarios in which the target speech can be different over time (such as in a conversation) and the masking sources can move. This could be done by simulating real-life scenarios through headphones so that rooms could be different and real, which has not been the case in our studies.

In order to address some of the model limitations and outlook listed in the above paragraphs, I am currently doing an additional study to use the model to predict individual psychometric functions for NH and HI listeners as well as percentage of correct words at fixed SNRs in 6 realistic, virtual-simulated environments involving unaided and aided listening with non-linear amplification. The overall level is known when predicting percentage of correct

words, which means that the assumption to approximate the overall level is no longer required. This ongoing study is ambitious but, if successful, the model will be able to predict speech intelligibility in aided listening for an individual HI listener in a realistic context.

References

J. B. Ahlstrom, A. R. Horwitz, and J. R. Dubno. Spatial separation benefit for unaided and aided listening. *Ear and Hearing*, 35(1):72–85, 2014. 12

M. A. Akeroyd. The psychoacoustics of binaural hearing. *International Journal of Audiology*, 45:25–33, 2006. 5

M. A. Akeroyd. Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, 47:53–71, 2008. 81

A. H. Andersen, J. M. de Haan, Z. Tan, and J. Jensen. Predicting the intelligibility of noisy and nonlinearly processed binaural speech. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(11):1908–1920, 2016. 22

A. H. Andersen, J. M. de Haan, Z. H. Tan, and J. Jensen. Refinement and validation of the binaural short time objective intelligibility measure for spatially diverse conditions. *Speech Communication*, 102:1–13, 2018. 14, 22, 23

ANSI S3.5. Methods for calculation of the speech intelligibility index. *American National Standards Institute, New York*, 1997. 14, 15, 31, 38, 47, 57, 68

T. L. Arbogast, C. R. Mason, and G. Kidd. The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 117(4):2169–2180, 2005. 11

J. Bench, A. Kowal, and J. Bamford. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, 13(3):108–112, 1979. 49, 66

R. A. Bentler and C. V. Pavlovic. Transfer functions and correction factors used in hearing aid evaluation and research free field to eardrum transfer function. *Ear and Hearing*, 10(1):58–63, 1989. 48

L. R. Bernstein and C. Trahiotis. Binaural signal detection, overall masking level, and masker interaural correlation: Revisiting the internal noise hypothesis. *The Journal of the Acoustical Society of America*, 124(6):3850–3860, 2008. 45, 46, 59, 60, 68, 77, 79, 119

L. R. Bernstein and C. Trahiotis. An interaural-correlation-based approach that accounts for a wide variety of binaural detection data. *The Journal of the Acoustical Society of America*, 141(2):1150–1160, 2017. 77, 80, 120

L. R. Bernstein and C. Trahiotis. Effects of interaural delay, center frequency, and no more than "slight" hearing loss on precision of binaural processing: Empirical data and quantitative modeling. *The Journal of the Acoustical Society of America*, 144(1):292–307, 2018. 77, 79

L. R. Bernstein and C. Trahiotis. Binaural detection as a joint function of masker bandwidth, masker interaural correlation, and interaural time delay: Empirical data and modeling. *The Journal of the Acoustical Society of America*, 148(6):3481–3488, 2020. 77

V. Best, E. R. Thompson, C. R. Mason, and G. Kidd. An energetic limit on spatial release from masking. *Journal of the Association for Research in Otolaryngology*, 14(4):603–610, 2013. 12

R. Beutelmann and T. Brand. Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 120(1):331–342, 2006. 7, 20, 46, 58, 60, 76, 77, 79, 80

R. Beutelmann, T. Brand, and B. Kollmeier. Prediction of binaural speech intelligibility with frequency-dependent interaural phase differences. *The Journal of the Acoustical Society of America*, 126(3):1359–1368, 2009. 42

R. Beutelmann, T. Brand, and B. Kollmeier. Revision, extension, and evaluation of a binaural speech intelligibility model. *The Journal of the Acoustical Society of America*, 127(4):2479–2497, 2010. 14, 20, 23, 32, 37, 39, 42, 46, 58, 60, 77, 80, 120

T. Biberger and S. D. Ewert. The role of short-time intensity and envelope power for speech intelligibility and psychoacoustic masking. *The Journal of the Acoustical Society of America*, 142(2):1098–1111, 2017. 14, 17, 23

N. Bisgaard, M. S. M. G. Vlaming, and M. Dahlquist. Standard audiograms for the IEC 60118-15 measurement procedure. *Trends in Amplification*, 14(2):113–120, 2010. xvii, 9

A. W. Bronkhorst and R. Plomp. The effect of head-induced interaural time and level differences on speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 83(4):1508–1516, 1988. 5, 7, 8, 63, 76

A. W. Bronkhorst and R. Plomp. Binaural speech intelligibility in noise for hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 86(4):1374–1383, 1989. 11, 63, 75, 76

A. W. Bronkhorst and R. Plomp. A clinical test for the assessment of binaural speech perception in noise. *International Journal of Audiology*, 29(5):275–285, 1990. 75, 76

I. C. Bruce, A. C. Léger, B. C. Moore, and C. Lorenzi. Physiological prediction of masking release for normal-hearing and hearing-impaired listeners. *Proceedings of Meetings on Acoustics*, 19 (1):050178, 2013. 61

D. S. Brungart and N. Iyer. Better-ear glimpsing efficiency with symmetrically-placed interfering talkers. *The Journal of the Acoustical Society of America*, 132(4):2545–2556, 2012. 6, 27, 29

S. Cameron and H. Dillon. Development of the listening in spatialized noise-sentences test (LISN-S). *Ear and Hearing*, 28:196–211, 4 2007. 66

A. Chabot-Leclerc, E. N. MacDonald, and T. Dau. Predicting binaural speech intelligibility using the signal-to-noise ratio in the envelope power spectrum domain. *The Journal of the Acoustical Society of America*, 140(1):192–205, 2016. 14, 18, 23

E. C. Cherry. Some experiments on the recognition of speech, with one and with 2 ears. *Journal of the Acoustical Society of America*, 25(5):975–979, 1953. 1

T. Y. C. Ching, H. Dillon, and D. Byrne. Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification. *The Journal of the Acoustical Society of America*, 103(2):1128–1140, 1998. 15, 58

B. Collin and M. Lavandier. Binaural speech intelligibility in rooms with variations in spatial location of sources and modulation depth of noise interferers. *The Journal of the Acoustical Society of America*, 134(2):1146–1159, 2013. xvii, xix, 4, 9, 14, 21, 24, 25, 27, 28, 30, 31, 32, 33, 39, 41, 42, 46, 47, 49, 54, 57, 60, 79, 93, 94, 95, 117, 118, 119

M. Cooke. A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, 119(3):1562–1573, 2006. 5, 19

S. Cosentino, T. Marquardt, D. McAlpine, J. F. Culling, and T. H. Falk. A model that predicts the binaural advantage to speech intelligibility from the mixed target and interferer signals. *The Journal of the Acoustical Society of America*, 135(2):796–807, 2014. 14, 21, 23

J. Cubick, J. M. Buchholz, V. Best, M. Lavandier, and T. Dau. Listening through hearing aids affects spatial perception and speech intelligibility in normal-hearing listeners. *The Journal of the Acoustical Society of America*, 144(5):2896–2905, 2018. 21, 30, 31, 32, 33, 39, 41, 47, 118

J. F. Culling and M. Lavandier. Binaural unmasking and spatial release from masking. In R. Y. Litovsky, M. J. Goupell, R. R. Fay, and A. N. Popper, editors, *Binaural Hearing*, pages 209–241. Springer International Publishing, 2021. 7

J. F. Culling and E. R. Mansell. Speech intelligibility among modulated and spatially distributed noise sources. *The Journal of the Acoustical Society of America*, 133(4):2254–2261, 2013. 6, 7, 8, 27, 28, 33, 37, 39, 118

J. F. Culling and Q. Summerfield. Measurements of the binaural temporal window using a detection task. *The Journal of the Acoustical Society of America*, 103(6):3540–3553, 1998. 27, 33, 37, 42

J. F. Culling, M. L. Hawley, and R. Y. Litovsky. The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources. *The Journal of the Acoustical Society of America*, 116(2), 2004. 5, 7, 8, 64, 68, 79, 121

J. F. Culling, M. L. Hawley, and R. Y. Litovsky. Erratum: The role head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources [J. Acoust. Soc. Am. 116, 1057 (2004)]. *The Journal of the Acoustical Society of America*, 118 (1):552, 2005. 19, 21, 31, 47, 64, 68, 79, 121

A. C. Davis and H. J. Hoffman. Hearing loss: Rising prevalence and impact. *Bulletin of the World Health Organization*, 97(10):646–646A, 2019. 1

J. G. Desloge, C. M. Reed, L. D. Braida, Z. D. Perez, and L. A. Delhorne. Speech reception by listeners with real and simulated hearing impairment: Effects of continuous and interrupted noise. *The Journal of the Acoustical Society of America*, 128(1):342–359, 2010. 11

H. Dillon. *Hearing Aids*. Thieme, New-York, 2nd edition edition, 2012. 12, 50, 67

P. L. Divenyi and K. M. Haupt. Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. I. Age and lateral asymmetry effects. *Ear and hearing*, 18(1):42–61, 1997. 11

P. L. Divenyi, P. B. Stark, and K. M. Haupt. Decline of speech understanding and auditory thresholds in the elderly. *The Journal of the Acoustical Society of America*, 118(2):1089–1100, 2005. 11

J. R. Dubno. Speech recognition across the life span: Longitudinal changes from middle-age to older adults. *American Journal of Audiology*, 24(2):84–87, 2015. 11

J. R. Dubno, J. B. Ahlstrom, and A. R. Horwitz. Spectral contributions to the benefit from spatial separation of speech and noise. *Journal of Speech, Language, and Hearing Research*, 45(6):1297–1310, 2002. 11

J. R. Dubno, J. B. Ahlstrom, and A. R. Horwitz. Binaural advantage for younger and older adults with normal hearing. *Journal of Speech, Language, and Hearing Research*, 51(2):539–556, 2008. 11

N. I. Durlach. Equalization and Cancellation Theory of Binaural Masking-Level Differences. *The Journal of the Acoustical Society of America*, 35(8):1206–1218, 1963. 6, 69

N. I. Durlach. Binaural signal detection: Equalization and cancellation theory. In J. Tobias, editor, *Foundations of Modern Auditory Theory*, volume II, pages 371–462. Academic, New York, 1972. 6, 31, 63, 64, 69, 76, 79, 116, 120, 121

N. I. Durlach, C. L. Thompson, and H. S. Colburn. Binaural interaction in impaired listeners: A review of past research. *International Journal of Audiology*, 20(3):181–211, 1981. 10, 63, 76

S. D. Ewert, W. Schubotz, T. Brand, and B. Kollmeier. Binaural masking release in symmetric listening conditions with spectro-temporally modulated maskers. *The Journal of the Acoustical Society of America*, 142(1):12–28, 2017. xviii, xxi, 6, 8, 29, 31, 34, 39, 40, 41, 43, 118

T. H. Falk, S. Cosentino, J. Santos, D. Suelzle, and V. Parsa. Non-intrusive objective speech quality and intelligibility prediction for hearing instruments in complex listening environments. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 7820–7824, 2013. 14, 16

J. M. Festen and R. Plomp. 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):1725–1736, 1990. 4, 47, 75, 93

P. J. Fitzgibbons and F. L. Wightman. Gap detection in normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 72(3):761–765, 1982. 10, 60

C. Füllgrabe, B. C. J. Moore, and M. A. Stone. Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Frontiers in Aging Neuroscience*, 6:347, 2015. 60

D. Fogerty, B. L. Carter, and E. W. Healy. Glimpsing speech in temporally and spectro-temporally modulated noise. *The Journal of the Acoustical Society of America*, 143(5):3047–3057, 2018. 4, 42

N. R. French and J. C. Steinberg. Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America*, 19(1):90–119, jan 1947. 3

C. Füllgrabe and B. C. Moore. Evaluation of a method for determining binaural sensitivity to temporal fine structure (tfs-af test) for older listeners with normal and impaired low-frequency hearing. *Trends in Hearing*, 21:1–14, 2017. 10, 60, 64, 73, 120

C. Füllgrabe and B. C. Moore. The Association Between the Processing of Binaural Temporal-Fine-Structure Information and Audiometric Threshold and Age: A Meta-Analysis. *Trends in Hearing*, 22:1–14, 2018. 73, 74

C. Füllgrabe, A. J. Harland, A. P. Sęk, and B. C. Moore. Development of a method for determining binaural sensitivity to temporal fine structure. *International Journal of Audiology*, 56 (12):926–935, 2017. 5, 65

GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, 396(10258):1204–1222, 2020. 1

S. A. Gelfand, L. Ross, and S. Miller. Sentence reception in noise from one versus two sources: Effects of aging and hearing loss. *The Journal of the Acoustical Society of America*, 83(1): 248–256, 1988. 11

E. L. J. George, J. M. Festen, and S. Theo Goverts. Effects of reverberation and masker fluctuations on binaural unmasking of speech. *The Journal of the Acoustical Society of America*, 132(3):1581–1591, 2012. 7, 63, 74, 76

B. R. Glasberg and B. C. J. Moore. Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *The Journal of the Acoustical Society of America*, 79(4): 1020–1033, 1986. 10, 16, 60

B. R. Glasberg, B. C. J. Moore, and S. P. Bacon. Gap detection and masking in hearing-impaired and normal-hearing subjects. *The Journal of the Acoustical Society of America*, 81 (5):1546–1556, 1987. 10

H. Glyde, L. Hickson, S. Cameron, and H. Dillon. Problems hearing in noise in older adults: a review of spatial processing disorder. *Trends in Amplification*, 15(3):116–126, 2011. 10, 11, 81

H. Glyde, J. M. Buchholz, H. Dillon, S. Cameron, and L. Hickson. The importance of interaural time differences and level differences in spatial release from masking. *The Journal of the Acoustical Society of America*, 134(2):EL147–EL152, 2013. 10, 117

H. Glyde, J. M. Buchholz, L. Nielsen, V. Best, H. Dillon, S. Cameron, and L. Hickson. Effect of audibility on spatial release from speech-on-speech masking. *The Journal of the Acoustical Society of America*, 138(5):3311–3319, 2015. 75

R. L. Goldsworthy and J. E. Greenberg. Analysis of speech-based speech transmission index methods with implications for nonlinear operations. *The Journal of the Acoustical Society of America*, 116(6):3679–3689, 2004. 13

S. T. Goverts and T. Houtgast. The binaural intelligibility level difference in hearing-impaired listeners: The role of supra-threshold deficits. *The Journal of the Acoustical Society of America*, 127(5):3073–3084, 2010. 7, 63, 74, 76

D. W. Grantham. Detectability of time-varying interaural correlation in narrow-band noise stimuli. *The Journal of the Acoustical Society of America*, 72(4):1178–1184, 1982. 27

D. W. Grantham and F. L. Wightman. Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation. *The Journal of the Acoustical Society of America*, 65(6):1509–1517, 1979. 27, 33, 37

C. F. Hauth and T. Brand. Modeling sluggishness in binaural unmasking of speech for maskers with time-varying interaural phase differences. *Trends in Hearing*, 22:1–10, 2018. 27, 33, 37, 47, 68

C. F. Hauth, N. Gößling, and T. Brand. Performance prediction of the binaural MVDR beamformer with partial noise estimation using a binaural speech intelligibility model. In *ITG-Fachbericht 282: Speech Communication*, pages 301–305, 2018. 20

C. F. Hauth, S. C. Berning, B. Kollmeier, and T. Brand. Modeling binaural unmasking of speech using a blind binaural processing stage. *Trends in Hearing*, 24:1–16, 2020. 14, 20, 23, 80, 81

M. L. Hawley, R. Y. Litovsky, and J. F. Culling. The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *The Journal of the Acoustical Society of America*, 115(2):833–843, 2004. 8

K. S. Helfer, G. R. Merchant, and P. A. Wasiuk. Age-related changes in objective and subjective speech perception in complex listening environments. *Journal of Speech, Language, and Hearing Research*, 60(10):3009–3018, oct 2017. 10, 11

ISO 3382-1. Acoustics — Measurement of room - Part 1: Performance spaces, 2009. (International Organization for Standardization, Geneva, Switzerland, 2001).

ISO 389-2. Acoustics - reference zero for the calibration of audiometric equipment - part 2: Reference equivalent threshold sound pressure levels for pure tones and insert earphones, 1994. (International Organization for Standardization, Geneva, Switzerland, 2001). 48

S. Jelfs, J. F. Culling, and M. Lavandier. Revision and validation of a binaural model for speech intelligibility in noise. *Hearing Research*, 275(1):96–104, 2011. 8, 30, 42

J. Jensen and C. H. Taal. An Algorithm for Predicting the Intelligibility of Speech Masked by Modulated Noise Maskers. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 24(11):2009–2022, 2016. 16

K. K. Jensen and J. G. Bernstein. The fluctuating-masker benefit for normal-hearing and hearing-impaired listeners with equal audibility at a fixed signal-to-noise ratio. *The Journal of the Acoustical Society of America*, 145(4):2113–2125, 2019. 10, 11, 12, 75

S.-H. Jin and P. B. Nelson. Speech perception in gated noise: The effects of temporal resolution. *The Journal of the Acoustical Society of America*, 119(5):3097–3108, 2006. 10

S. Jørgensen and T. Dau. Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. *The Journal of the Acoustical Society of America*, 130(3):1475–1487, 2011. 16

S. Jørgensen, S. D. Ewert, and T. Dau. A multi-resolution envelope-power based model for speech intelligibility. *The Journal of the Acoustical Society of America*, 134(1):436–446, 2013. 14, 16, 18

A. Josupeit and V. Hohmann. Modeling speech localization, talker identification, and word recognition in a multi-talker setting. *The Journal of the Acoustical Society of America*, 142 (1):35–54, jul 2017. 14, 19, 23, 80

C. Kamm, D. D. Dirks, and M. R. Mickey. Effect of sensorineural hearing loss on loudness discomfort level and most comfortable loudness judgments. *Journal of Speech and Hearing Research*, 21(4):668–681, 1978. 10

J. M. Kates and K. H. Arehart. Coherence and the speech intelligibility index. *The Journal of the Acoustical Society of America*, 117(4):2224–2237, 2005. 12, 14, 15

J. M. Kates and K. H. Arehart. The hearing-aid speech perception index (HASPI). *Speech Communication*, 65:75–93, 2014. 14, 17, 81

M. Kathleen Pichora-Fuller, B. A. Schneider, and M. Daneman. How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, 97(1): 593–608, 1995. 10

G. Keidser, H. Dillon, M. Flax, T. Ching, and S. Brewer. The NAL-NL2 prescription procedure. *Audiology Research*, 1(1S):1–3, 2011. 12

G. Kidd and H. S. Colburn. Informational masking in speech recognition. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay, editors, *The Auditory System at the Cocktail Party*, pages 75–109. Springer International Publishing, Cham, 2017. 60

S. H. Kim, R. D. Frisina, and D. R. Frisina. Effects of age on speech understanding in normal hearing listeners: Relationship between the auditory efferent system and speech intelligibility in noise. *Speech Communication*, 48(7):855–862, 2006. 11

A. King, K. Hopkins, and C. J. Plack. The effects of age and hearing loss on interaural phase difference discrimination. *The Journal of the Acoustical Society of America*, 135(1):342–351, 2014. 10, 60, 64

U. Kjems, J. B. Boldt, M. S. Pedersen, T. Lunner, and D. Wang. Role of mask pattern in intelligibility of ideal binary-masked noisy speech Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The Journal of the Acoustical Society of America*, 126(3):1415–1426, 2009. 12

A. J. Kolarik and J. F. Culling. Measurement of the binaural auditory filter using a detection task. *The Journal of the Acoustical Society of America*, 127(5):3009–3017, 2010. 42

B. Kollmeier and J. Kiessling. Functionality of hearing aids: state-of-the-art and future model-based solutions. *International Journal of Audiology*, 57:S3–S28, 2018. 12

K. D. Kryter. Methods for the calculation and use of the articulation index. *The Journal of the Acoustical Society of America*, 34(11):1689, 1962. 15, 38

M. Lavandier and V. Best. Modeling binaural speech understanding in complex situations. In J. Blauert and J. Braasch, editors, *The technology of binaural understanding, chapter 19*. Springer, Berlin–Heidelberg–New York NY, 2020. 18, 19

M. Lavandier and J. F. Culling. Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer. *The Journal of the Acoustical Society of America*, 123(4):2237–2248, 2008. 9, 21, 60

M. Lavandier and J. F. Culling. Prediction of binaural speech intelligibility against noise in rooms. *The Journal of the Acoustical Society of America*, 127(1):387–399, 2010. 7, 9, 21, 30, 42, 60, 63, 64, 66, 75, 76, 81

M. Lavandier, S. Jelfs, J. F. Culling, A. J. Watkins, A. P. Raimond, and S. J. Makin. Binaural prediction of speech intelligibility in reverberant rooms with multiple noise sources. *The Journal of the Acoustical Society of America*, 131(1):218–231, 2012. xix, 8, 9, 21, 30, 42, 46, 47, 49, 53, 56, 60, 93, 95, 119

M. Lavandier, J. M. Buchholz, and B. Rana. A binaural model predicting speech intelligibility in the presence of stationary noise and noise-vocoded speech interferers for normal-hearing and hearing-impaired listeners. *Acta Acustica united with Acustica*, 104(5):909–913, 2018. xvi, xvii, 14, 21, 23, 24, 25, 43, 45, 46, 47, 48, 49, 54, 55, 59, 79, 97, 98, 100, 117, 119, 121

T. Leclère, M. Lavandier, and J. F. Culling. Speech intelligibility prediction in reverberation: Towards an integrated model of speech transmission, spatial unmasking, and binaural de-reverberation. *The Journal of the Acoustical Society of America*, 137(6):3335–3345, 2015. 9, 14, 21, 60, 66, 81

N. A. Lesica. Why do hearing aids fail to restore normal auditory perception? *Trends in Neurosciences*, 41(4):174–185, 2018. 10

J. C. R. Licklider. The influence of interaural phase relations upon the masking of speech by white noise the influence of interaural phase on interaural summation and inhibition. *The Journal of the Acoustical Society of America*, 20(2):150–159, 1948. 7, 76

D. McFadden. Masking-level differences determined with and without interaural disparities in masker intensity. *The Journal of the Acoustical Society of America*, 44(1):212–223, 1968. 45

R. M. Meyer and T. Brand. Comparison of different short-term speech intelligibility index procedures in fluctuating noise for listeners with normal and impaired hearing. *Acta Acustica united with Acustica*, 99(3):442–456, 2013. 15

J. Mi and S. H. Colburn. A binaural grouping model for predicting speech intelligibility in multitalker environments. *Trends in Hearing*, 20:1–12, 2016. 14, 19, 23, 80

G. A. Miller and J. C. R. Licklider. The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America*, 22(2):167–173, 1950. 4

B. C. Moore. The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology*, 9(4):399–406, 2008. 10

B. C. Moore, B. R. Glasberg, C. J. Plack, and a. K. Biswas. The shape of the ear's temporal window. *The Journal of the Acoustical Society of America*, 83(3):1102–1116, 1988. 27, 39, 42

B. C. Moore, B. R. Glasberg, and M. A. Stone. Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: CAMEQ2-HF. *International Journal of Audiology*, 49(3):216–227, 2010. 12

B. C. J. Moore and B. R. Glasberg. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America*, 74(3):750–753, 1983. 15, 31, 66

B. C. J. Moore and B. R. Glasberg. A revised model of loudness perception applied to cochlear hearing loss. *Hearing Research*, 188:70–88, 2004. 48, 61, 119

B. C. J. Moore, M. Wojtczak, and D. A. Vickers. Effect of loudness recruitment on the perception of amplitude modulation. *The Journal of the Acoustical Society of America*, 100(1):481–489, 1996. 10

B. C. J. Moore, B. R. Glasberg, and D. A. Vickers. Further evaluation of a model of loudness perception applied to cochlear hearing loss. *The Journal of the Acoustical Society of America*, 106(2):898–907, 1999. 10

D. R. Murphy, M. Daneman, and B. A. Schneider. Why do older adults have difficulty following conversations? *Psychology and Aging*, 21(1):49–61, 2006. 11

T. Neher, S. Laugesen, N. Søgaard Jensen, and L. Kragelund. Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities? *The Journal of the Acoustical Society of America*, 130(3):1542–1558, 2011. 10, 60, 64, 73, 74

T. Neher, T. Lunner, K. Hopkins, and B. C. J. Moore. Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners (L). *The Journal of the Acoustical Society of America*, 131(4):2561–2564, 2012. 60, 64, 73, 74

O. Nordvik, P. O. Laugen Heggdal, J. Brännström, F. Vassbotn, A. K. Aarstad, and H. J. Aarstad. Generic quality of life in persons with hearing loss: A systematic literature review. *BMC Ear, Nose and Throat Disorders*, 18(1):1–13, jan 2018. 1

W. Passchier-Vermeer. Hearing loss due to continuous exposure to steady-state broad-band noise. *Journal of the Acoustical Society of America*, 56(5):1585–1593, 1974. 9

R. D. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice. An efficient auditory filterbank based on the gammatone function. presented to the Institute of Acoustics speech group on auditory modelling at the Royal Signal Research Establishment., 1987. 31

J. Peissig and B. Kollmeier. Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. *The Journal of the Acoustical Society of America*, 101(3):1660–1670, 1997. xvii, 7, 8

S. A. Phatak and K. W. Grant. Phoneme recognition in modulated maskers by normal-hearing and aided hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 132 (3):1646–1654, 2012. 10

S. A. Phatak and K. W. Grant. Phoneme recognition in vocoded maskers by normal-hearing and aided hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 136 (2):859–866, 2014. 10

I. Pieper, M. Mauermann, D. Oetting, B. Kollmeier, and S. D. Ewert. Physiologically motivated individual loudness model for normal hearing and hearing impaired listeners. *The Journal of the Acoustical Society of America*, 144(2):917–930, 2018. 49, 61

C. J. Plack and B. C. J. Moore. Temporal window shape as a function of frequency and level. *The Journal of the Acoustical Society of America*, 87(5):2178–2187, 1990. 27, 39, 42

R. Plomp. Auditory handicap of hearing impairment and the limited benefit of hearing aids. *The Journal of the Acoustical Society of America*, 63(2):533–549, 1978. 46, 58

B. Rana and J. M. Buchholz. Better-ear glimpsing at low frequencies in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 140(2):1192–1205, 2016. xviii, xx, 46, 47, 48, 49, 50, 51, 52, 55, 60, 66, 76, 97, 98, 99

B. Rana and J. M. Buchholz. Effect of audibility on better-ear glimpsing as a function of frequency in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 143(4):2195–2206, 2018a. xix, xx, 12, 46, 47, 48, 49, 50, 52, 55, 57, 60, 61, 97, 99

B. Rana and J. M. Buchholz. Effect of improving audibility on better-ear glimpsing using nonlinear amplification. *The Journal of the Acoustical Society of America*, 144(6):3465–3474, 2018b. xix, xx, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 60, 97, 98, 99, 100

H. Relaño-Iborra, T. May, J. Zaar, C. Scheidiger, and T. Dau. Predicting speech intelligibility based on a correlation metric in the envelope power spectrum domain. *The Journal of the Acoustical Society of America*, 140(4):2670–2679, 2016. 17

H. Relaño-Iborra, J. Zaar, and T. Dau. A speech-based computational auditory signal processing and perception model. *The Journal of the Acoustical Society of America*, 146(5):3306–3317, 2019. 14, 17

J. Rennies, T. Brand, and B. Kollmeier. Prediction of the influence of reverberation on binaural speech intelligibility in noise and in quiet. *The Journal of the Acoustical Society of America*, 130(5):2999–3012, 2011. 9, 14, 20

K. S. Rhebergen and N. J. Versfeld. A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *The Journal of the Acoustical Society of America*, 117(4):2181–2192, 2005. 15, 30, 39, 42

K. S. Rhebergen, N. J. Versfeld, and W. A. Dreschler. Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise. *The Journal of the Acoustical Society of America*, 120(6):3988–3997, 2006. 15, 42

K. S. Rhebergen, J. Lyzenga, W. A. Dreschler, and J. M. Festen. Modeling speech intelligibility in quiet and noise in listeners with normal and impaired hearing. *The Journal of the Acoustical Society of America*, 127(3):1570–1583, 2010. 15

D. E. Robinson and L. A. Jeffress. Effect of Varying the Interaural Noise Correlation on the Detectability of Tonal Signals. *The Journal of the Acoustical Society of America*, 35(12): 1947–1952, 1963. 7, 45, 46

A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010. 31, 79

S. Santurette and T. Dau. Relating binaural pitch perception to the individual listener's auditory profile. *The Journal of the Acoustical Society of America*, 131(4):2968–2986, 2012. 60, 64, 73, 74

M. R. Schädler, A. Warzybok, and B. Kollmeier. Objective prediction of hearing aid benefit across listener groups using machine learning: Speech recognition performance with binaural noise-reduction algorithms. *Trends in Hearing*, 22:233121651876895, 2018. 14, 22, 23, 80

C. Scheidiger, L. H. Carney, T. Dau, and J. Zaar. Predicting speech intelligibility based on across-frequency contrast in simulated auditory-nerve fluctuations. *Acta Acustica united with Acustica*, 104(5):914–917, 2018. 61

B. A. Schneider and M. K. Pichora-Fuller. Age-related changes in temporal processing: Implications for speech perception. *Seminars in Hearing*, 22(3):227–238, 2001. 10, 60

K. Smeds, F. Wolters, and M. Rung. Estimation of signal-to-noise ratios in realistic sound scenarios. *Journal of the American Academy of Audiology*, 26:183–196, 2015. 61, 81

P. E. Souza, K. H. Arehart, J. M. Kates, N. B. H. Croghan, and N. Gehani. Exploring the limits of frequency lowering. *Journal of Speech, Language, and Hearing Research*, 56(5):1349–1363, 2013. 12

A. Spoor. Presbycusis values in relation to noise induced hearing loss. *International Journal of Audiology*, 6(1):48–57, 1967. 9

H. J. M. Steeneken and T. Houtgast. A physical method for measuring speech-transmission quality. *The Journal of the Acoustical Society of America*, 67(1):318–326, 1980. 13, 14, 38

O. Strelcyk and T. Dau. Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *The Journal of the Acoustical Society of America*, 125(5):3328–3345, 2009. 10, 60, 64, 73, 74

G. A. Studebaker and R. L. Sherbecoe. Intensity-importance functions for bandlimited monosyllabic words. *The Journal of the Acoustical Society of America*, 111(3):1422–1436, 2002. 39

C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen. An algorithm for intelligibility prediction of time-frequency weighted noisy speech. *IEEE Transactions on Audio, Speech and Language Processing*, 19(7):2125–2136, 2011. 14, 16

Y. Tang, M. Cooke, B. M. Fazenda, and T. J. Cox. A metric for predicting binaural speech intelligibility in stationary noise and competing speech maskers. *The Journal of the Acoustical Society of America*, 140(3):1858–1870, sep 2016. 14, 19

S. J. Van Wijngaarden and R. Drullman. Binaural intelligibility prediction based on the speech transmission index. *The Journal of the Acoustical Society of America*, 123(6):4514–4523, 2008. 14, 18

T. Vicente and M. Lavandier. Further validation of a binaural model predicting speech intelligibility against envelope-modulated noises. *Hearing Research*, 390:107937, 2020. xviii, 27, 46, 47, 49, 68, 118

T. Vicente, M. Lavandier, and J. M. Buchholz. A binaural model implementing an internal noise to predict the effect of hearing impairment on speech intelligibility in non-stationary noises. *The Journal of the Acoustical Society of America*, 148(5):3305–3317, 2020. 45, 63, 64, 68, 75, 76, 78, 119

H. vom Hömel. *Zur Bedeutung der Übertragungseigenschaften des Außenohrs sowie des binauralen Hörsystems bei gestörter Sprachübertragung (On the importance of the transmission properties of the outer ear and the binaural auditory system in disturbed speech transmisson).* PhD thesis, RWTH, Aachen, 1984. 77

R. Wan, N. I. Durlach, and H. S. Colburn. Application of an extended equalization-cancellation model to speech intelligibility with spatially distributed maskers. *The Journal of the Acoustical Society of America*, 128(6):3678–3690, 2010. 18, 69

R. Wan, N. I. Durlach, and H. S. Colburn. Application of a short-time version of the Equalization-Cancellation model to speech intelligibility experiments with speech maskers. *The Journal of the Acoustical Society of America*, 136(2):768–776, 2014. 14, 18, 23, 42

A. Weisser and J. M. Buchholz. Conversational speech levels and signal-to-noise ratios in realistic acoustic conditions. *The Journal of the Acoustical Society of America*, 145(1):349–360, 2019. 57

World Health Organization. *Global costs of unaddressed hearing loss and cost-effectiveness of interventions: a WHO report, 2017.* World Health Organization, 2017. 1

U. T. Zwicker and E. Zwicker. Binaural masking-level difference as a function of masker and test-signal duration. *Hearing Research*, 13:215–219, 1984. 63

A Description of the experiment (VL) conducted in Chapter III

Aim

VL employed modulated maskers, different levels of reverberation, in an asymmetrical configuration. It also aimed at emphasizing the contribution of better ear listening to speech intelligibility. Hence, stimuli with and without ITDs were considered using BRIRs and spectral-envelope impulse responses (SEIRs, Lavandier et al., 2012). SEIRs were obtained by removing the ITDs and reverberation tails of the BRIRs, while preserving their long-term spectrum (at each ear). In the following, BRIRs and SEIRs are associated with the "ITD+ILD" and "no ITD/no tail" conditions, respectively. Because the stimuli in the no ITD/no tail condition did not contain reverberation tails, the influence of reverberation filling in the masker modulation gaps was varied here in an asymmetrical condition.

It was hypothesized that, for the conditions with SRM, higher SRTs should be obtained in the no ITD/no tail condition compared to the ITD+ILD condition due to the absence of ITDs/binaural unmasking. It was also hypothesized that the difference between SRTs measured with steady-state and modulated noises should be larger in the no ITD/no tail conditions, at least at large distances, when reverberation tails fill in the dips of the modulated noise so that it becomes steady-state.

Stimuli and apparatus

The stimuli were produced as done by Collin and Lavandier (2013). A male speaker uttered semantically unpredictable sentences in French that contained four key words. The anechoic recordings were used as the basis of all stimuli. The maskers were noises (SSNs) either steady-state or modulated by an envelope extracted from a speech signal (1-voice modulated noises). A long steady-state noise was obtained by concatenating several lists of sentences, taking the Fourier transform of the resulting signal, randomizing its phase, and finally taking its inverse Fourier transform.

To create the speech-modulated noises, the envelopes of the sentences were extracted as proposed by Festen and Plomp (1990), then concatenated by pairs keeping a 100-ms silence between them. The modulated noises were obtained by multiplying these envelopes with the steady-state noises. During the test, a masker envelope was never the same as the target envelope.

Real-room listening was simulated over headphones by convolving the anechoic stimuli with the BRIRs. These BRIRs were measured by Lavandier et al. (2012) in a meeting room (meeting room 1). SEIRs were also used in order to evaluate the contributions of reverberation tails and binaural unmasking. SEIRs were designed to remove the ITDs and reverberation tails of the BRIRs, but preserve room coloration and long-term ILDs when present, since SEIRs retain the same long-term spectrum as their corresponding BRIR. Full information about the measurements and processing can be found in Lavandier et al. (2012).

The convolution by a BRIR can introduce level differences in the resulting signals across different positions. To avoid these level effects, the left-right average of the RMS power of the convolved stimuli was equalized before the experiment, i.e., the levels of the spatialized stimuli were equalized at the ears of the listeners while preserving the ILDs.

Signals were digitally mixed, D/A converted, and amplified using a Lynx TWO sound card. They were presented to listeners over Sennheiser HD 650 headphones in a double-walled sound-proof booth. A computer screen was visible outside the booth window. A keyboard was inside the booth to gather the transcripts.

Design

The target was simulated at 0.65 m and +25° from the listener. Two types of noise (steady-state or 1-voice modulated) were tested at two distances, 0.65 m and 5 m referred to as "Near" and "Far". Two noise azimuths were also tested, one identical (+25°) and one different (-25°) from the one of the target (+25°). The target (+25°/Near) was presented against a single noise in each condition. The combination of all these experimental factors resulted in 16 conditions (NOISE MODULATION{steady-state, modulated} x NOISE DISTANCE{near, far} x NOISE AZIMUTH{= target, ≠ target} x IMPULSE RESPONSE TYPE{itd+ild, no itd/no tail}).

Procedure

The adaptive procedure used to measure the SRTs was similar to the one used by Collin and Lavandier (2013), except that Collin and Lavandier varied the target level and kept the noise level constant to control the SNR, but the overall sound level varied during the measurements. In the current experiment the overall level was fixed at 70 dB SPL (calibrated using a MK2/NCF1 dummy head, Neutrik Cortex Instrument), and instead of applying formula 1 in Collin and Lavandier (2013) to the target level as they did, it was applied here on the SNR.

The results of a listener were discarded from the data if there was no inversion in the adaptive procedure to measure a SRT. It occurred only once during the experiment. Another listener was enrolled to substitute the participant whose results were discarded.

Listeners

Seventeen French native speakers participated in the experiment. The data of one participant was discarded because one SRT measurement failed (see previous section). All participants had an hearing threshold equal to or better than 20 dB HL from 125 Hz to 8 000 Hz. None of them was familiar with the speech material. All provided written informed consent and were paid for their participation.

Results

Figure III-6 presents the SRTs measured in VL, averaged across listeners and plotted as a function of the noise position, with one panel for each type of noise modulation. There was no difference between the no ITD/no tail and ILD+ITD conditions for the steady-state noise. For the modulated noise, the SRTs were lower for the no ITD/no tail condition than for the ITD+ILD condition. For both types of noise modulation and impulse response, when the masker was spatially separated from the target, listeners had SRMs of at least 4 dB. In the "far" conditions, SRMs were lower but at least 2 dB.

A repeated-measure analysis of variance (ANOVA) confirmed significant effects of the impulse response type [$F(1,15) = 15.5$, $p = 0.001$], the noise distance [$F(1,15) = 35.0$, $p < 0.0001$], the noise modulation [$F(1,15) = 238.4$, $p < 0.0001$] and the noise azimuth [$F(1,15) = 195.4$, $p < 0.0001$]. Two interactions were significant, between the impulse response type and the noise modulation [$F(1,15) = 38.8$, $p < 0.0001$] and between the noise distance and azimuth [$F(1,15) = 188.6$, $p < 0.0001$].

Discussion

SRTs in the modulated noise are consistently lower than those obtained in steady-state noise, i.e., listeners benefited from the masker gaps. A 5-dB SRM can be observed at near distance when the masker was moved from the =target-near position (co-located configuration) to the ≠target-near position. The SRM was reduced to about 2 dB in the far conditions (resulting in the interaction of the effects of masker distance and azimuth). This can be explained by

the increased effect of reverberation on the masker (Lavandier et al., 2012), which impairs both better-ear listening (by reducing head-shadow) and binaural unmasking (by decorrelating the masker at the two ears). SRTs were on average lower in the far conditions compared to the near conditions, probably highlighting a beneficial effect of room coloration in this particular configuration, as already observed previously (Lavandier et al., 2012; Collin and Lavandier, 2013).

Two effects associated with the no ITD/no tail condition may account for the interaction of the effects of noise modulation and impulse response. Stimuli without ITDs prevent listeners to use binaural unmasking, hence impair intelligibility. Under no ITD/no tail conditions reverberation does not fill in the masker dips. This allows listeners to use dip listening, thus enhancing speech intelligibility. These two counteracting effects could explain the difference between the ILD+ITD and the no ITD/no tail data for the modulated maskers in the bottom panel of Fig. III-6 (black symbols vs. grey symbols). SRTs are lower in the no ITD/no tail condition compared to the ILD+ITD condition, suggesting that the positive effect of having no reverberation tails is stronger that the negative effect of removing binaural unmasking. The difference is about 1 dB for the near distance and 2 dB for the far distance. The 1-dB intelligibility enhancement, observed when the masker distance increased, is consistent with the following explanation: when the masker is further away from the listener, it has more energy in its reverberation tails. It fills the masker gaps more, hence triggers a larger difference than under the (no ITD/no tail) SEIR conditions.

With the steady-state noise, no effect of the reverberation tails was expected. The difference between ITD+ILD and no ITD/no tail conditions should be limited to the involvement of binaural unmasking in the ITD+ILD condition. This should have led to lower SRTs compared to the no ITD/no tail. No significant effect of binaural unmasking was observed here at near distance even if it was previously observed using the same impulse responses (Lavandier et al., 2012). That effect was limited to about 1 dB, which could explain its lack of significance in the present experiment.

B

Comparison of the predictions from the models proposed by Lavandier et al. (2018) with the predictions from the model proposed in Chapter IV

This document presents a detailed comparison between the predictions of the model proposed in the present study with those of the normal-hearing (NH) and hearing-impaired (HI) models described in Lavandier et al. (2018, Lav18 models). Here, the Lav18 models (as well as the proposed model) convert binaural ratios into predict SRTs by considering both groups simultaneously (see Sec. IV.2.3 in the paper). The predictions of the proposed model are replotted in the below figures to ease the comparison.

Experiment of Rana and Buchholz (2016)

Figure B-1 shows the mean SRTs measured by Rana and Buchholz (2016) as a function of the masker type. The co-located and separated conditions are plotted as open and filled symbols, respectively, while grey symbols refer to the data and predictions derived for the HI listeners and the black symbols those for the NH listeners. The data are represented as circles, the predictions of the proposed model as downward triangles and those of the Lav18 models as squares. The Lav18 models are not able to predict well the difference between HI and NH listeners. The average difference in measured SRTs across conditions between NH and HI listeners is about 7.2 dB, while the Lav18 models predict only 3.3 dB (in comparison, the proposed model predicts 6.4 dB). This degrades the Lav18 models overall performances (compared to the proposed model) as observed in Table B.a.

Experiment of Rana and Buchholz (2018a)

Figure B-2 shows the SRTs as a function of the masker sensation level measured by Rana and Buchholz (2018a). The mean measured SRTs are shown as open (co-located) and filled (separated) circles, the mean predicted SRTs of the proposed model as dashed (co-located) and solid (separated) lines, and the mean predicted SRTs of the Lav18 models as open (co-located) and filled (separated) squares. The left panel shows the SRTs derived for the NH listeners and the right panel those for the HI listeners. As a general observation, the Lav18 models tend to overestimate intelligibility for the HI listeners, i.e., on the right panel the squares are below circles (except at 30 dB SL). Predictions of the Lav18 model for the NH listeners are as good as for the proposed model except at 0 dB SL, where it overestimates the SRTs by about 2.5 dB. The performance statistics of the Lav18 models are just slightly worse than those of the proposed model (Table B.a).

Experiment of Rana and Buchholz (2018b)

Figure B-3 presents the SRTs measured by Rana and Buchholz (2018b) as a function of the spatialization method. The data and predictions from the proposed and Lav18 models are shown as circles, downward triangles and squares, respectively, in grey for the HI listeners and in black for the NH listeners. The Lav18 models predict about 0.3 dB of difference across spatial

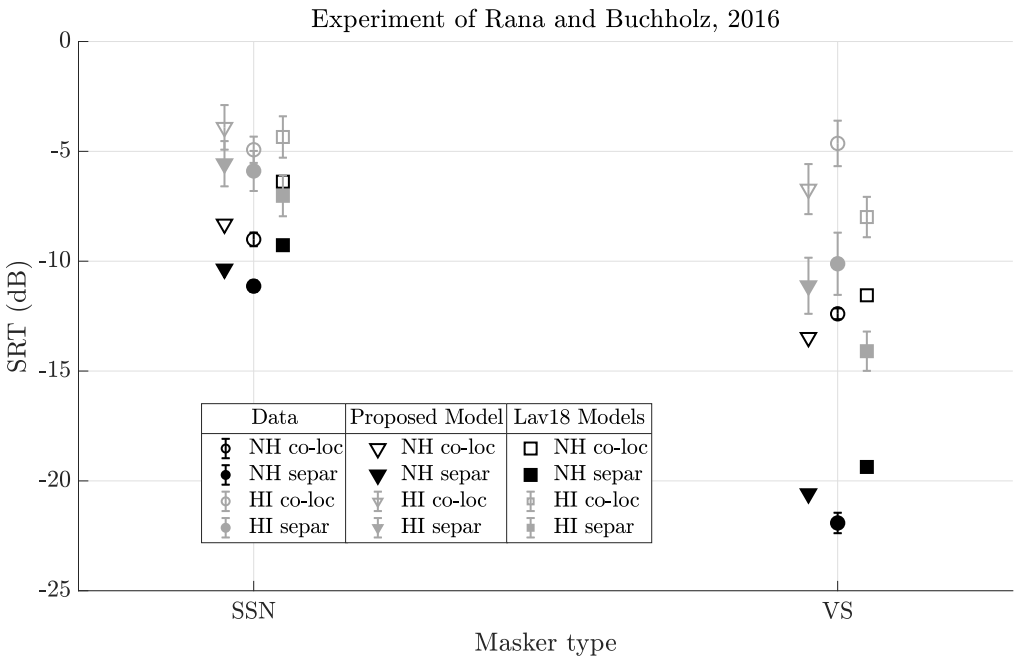


FIGURE B-1: Mean SRTs with ±1 standard errors across NH listeners (black circles) and HI listeners (grey circles) measured by Rana and Buchholz (2016) as a function of masker type (VS or SSN). The two maskers were either co-located with the target in front of the listener ("co-loc", open symbols) or simulated on each side of the listener at +/- 90° ("separ", filled symbols). Mean predicted SRTs are displayed as downward triangles (proposed model) and squares (Lav18 models) with the same filling and color patterns as the data.

conditions between the HI and NH listeners (difference between grey and black squares), which is about 5 dB less than what is observed in the data (difference between grey and black circles). The proposed model better predicts the difference between NH and HI listeners (difference between grey and black downward triangles), even though it is still underestimated by about 2 dB.

Summary of the model performances across experiments

Table B.a presents the performances of the Lav18 and proposed models for each experiment. Compared to the Lav18 models, the model proposed in this study improves the correlations (r and r_s) and reduces the mean and maximum absolute errors ($MeanErr$ and $MaxErr$, respectively) between data and predictions. The discrepancy of the Lav18 predictions for the experiments of Rana and Buchholz (2016, 2018b) are quantified by a MeanErr at least 1.1 dB higher than the one computed with the proposed model. The correlation for the experiment of Rana and Buchholz (2018b) is improved from 0.71 to 0.97 with the proposed model.

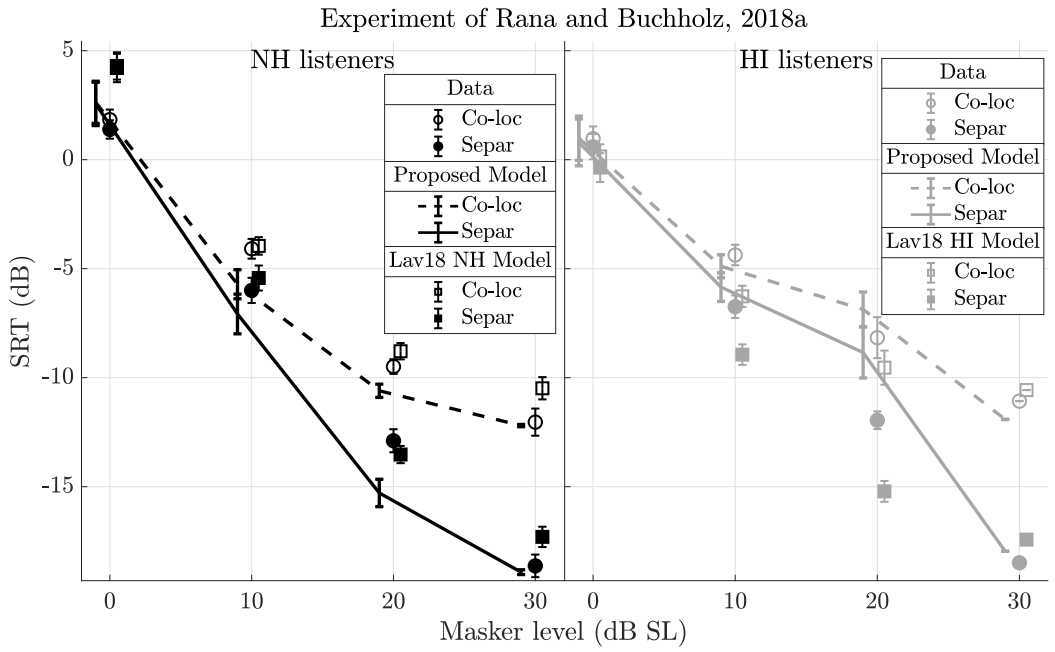


FIGURE B-2: Mean SRTs (circles) with ±1 standard errors across NH listeners (left panel) and HI listeners (right panel) measured by Rana and Buchholz (2018a) at four overall masker sensation levels. Predicted SRTs are plotted with lines for the proposed model and with squares for the Lav18 models. The two maskers were VSs either co-located with the target in front of the listener ("co-loc", open circles, open squares and dashed lines for data, Lav18 models and proposed model, respectively) or simulated on each side of the listener at +/- 90° ("separ", filled circles, filled squares and solid lines for data, Lav18 models and proposed model, respectively).

Experiment	r Lav18; Proposed	r_s Lav18; Proposed	$MeanErr$ Lav18; Proposed	$MaxErr$ Lav18; Proposed
Rana and Buchholz (2016)	0.89; 0.98	0.76; 0.90	2.1; 1.0	4.0; 2.1
Rana and Buchholz (2018a)	0.97; 0.98	0.97; 0.98	1.4; 1.0	3.3; 3.1
Rana and Buchholz (2018b)	0.71; 0.97	0.79; 0.98	2.7; 1.1	3.6; 2.4

TABLE B.A: Comparison of performances between the Lav18 models and the model proposed in Chapter IV.

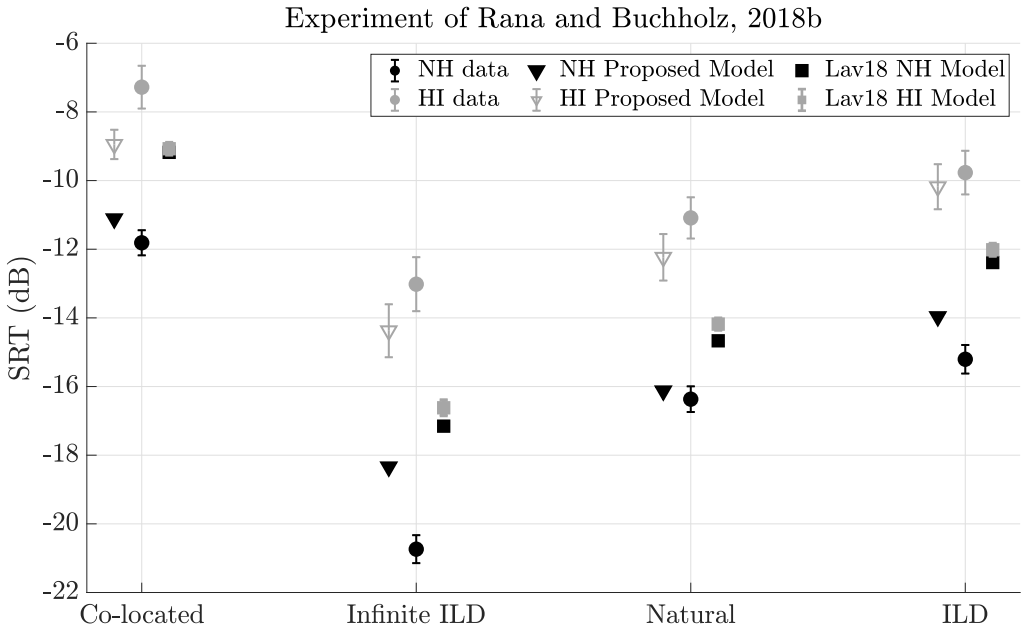


FIGURE B-3: Mean SRTs with ± 1 standard errors across NH listeners (in black circles) and HI listeners (in grey circles) measured by Rana and Buchholz (2018b) as a function of spatialization method. The two maskers were VS. Predicted SRTs obtained with the proposed model and Lav18 models are displayed as downward triangles and squares, respectively, with the same color pattern as the data.

C

Comparison between the model developed in Chapter IV and the model with the revised jitters proposed in Chapter V

The revised jitters that were applied to the Vic20 model were designed to account for the effect of low presentation levels on binaural unmasking, as observed in the SRT data of subset II (Table V.a). In the following, the effect of the original and revised jitters on model predictions are compared to quantify the advantage of audibility on binaural unmasking, associated with increasing the presentation level of the stimuli and computed as the SRTs at 40 dB SPL minus those at 50 and 60 dB SPL. This relative measure is introduced because it focuses on the improvement in the predictions provided by the revised jitters.

Figure C-1 shows the measured and predicted audibility advantages. As mentioned in the paper, the Vic20 model predicts well the data collected with HI listeners, however, the effect of presentation level is not predicted for the NH listeners.

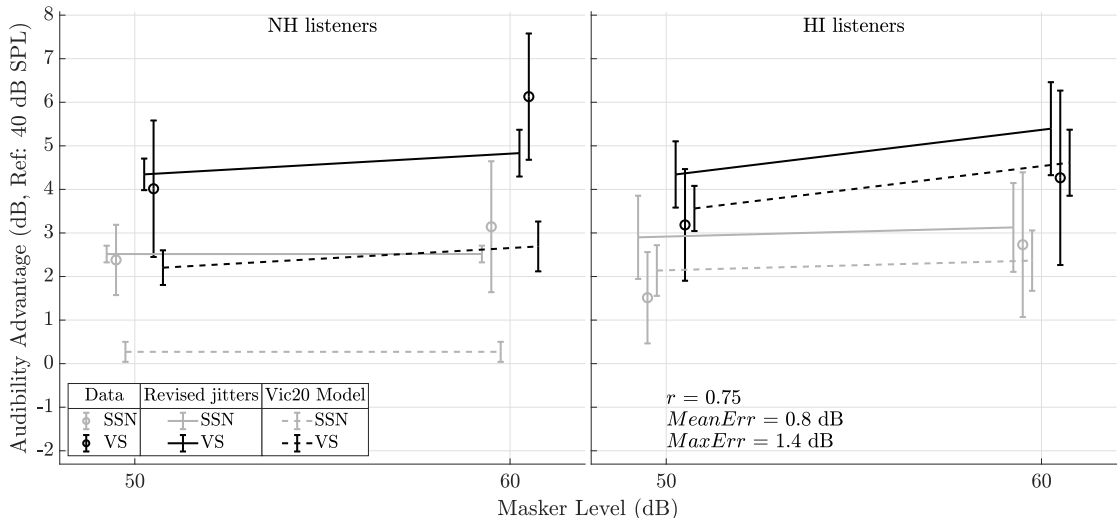


FIGURE C-1: Same as Fig. V-5 but for audibility advantage obtained by subtracting the SRTs at 50 and 60 dB SPL from the SRTs at 40 dB SPL.

The level-dependent jitters provide more accurate predictions as quantified by better performance statistics (see Table C.a). Computing the performances across all conditions, r and $MeanErr$ are only slightly improved with the revised jitters but $MaxErr$ decreases from 3.1 to 2 dB. Concerning the conditions at 60 dB SPL and the binaural unmasking advantage, the performance statistics of the two models are equal. When considering the anechoic conditions at different presentation levels, the revised jitters increase r from 0.85 to 0.93 and decrease $MaxErr$ from 3.1 to 2.0 dB. The major improvements are observed when predicting the audibility advantages. Thanks to the revised jitters r increases from 0.39 to 0.75, $MeanErr$ decreases from

Data	r Vic20; Rev.	$MeanErr$ (dB) Vic20; Rev.	$MaxErr$ (dB) Vic20; Rev.
All	0.91; 0.92	1.0; 0.9	3.1; 2.0
Conditions at 60 dB SPL	0.92; 0.92	0.9; 0.9	2.2; 2.0
BU_{Adv}	0.92; 0.92	0.5; 0.5	1.1; 1.1
Anechoic conditions	0.85; 0.93	1.1; 1.1	3.1; 2.0
Aud_{Adv}	0.39; 0.75	1.5; 0.8	3.4; 1.4

Table C.a: Performance statistics of the Vic20 model and the model with revised jitters (Rev.). The row "All" reports the performance statistics computed with the predicted SRTs across all conditions. $\mathrm{BU_{Adv}}$ and $\mathrm{Aud_{Adv}}$ stand for binaural unmasking advantage and audibility advantage, respectively.

1.5 to 0.8 dB and $MaxErr$ decreases from 3.4 to 1.4 dB.

D
Ethic approval letter

Human Sciences Subcommittee
Macquarie University, North Ryde
NSW 2109, Australia

MACQUARIE University

14/08/2019

Dear Associate Professor Buchholz,

Reference No: 5201955629923
Project ID: 5562
Title: Effect of interaural time difference sensitivity on speech intelligibility in normal-hearing and hearing-impaired listeners

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 Associate Professor Joerg Buchholz, and other personnel: Mr Thibault Vicente.

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

The Faculty Ethics Subcommittees at Macquarie University operate in accordance with the National Statement on Ethical Conduct in Human Research 2007, (updated July 2018), [Section 5.2.22].

E
Authorship contribution statement forms

MACQUARIE UNIVERSITY
AUTHORSHIP CONTRIBUTION STATEMENT

In accordance with the *Macquarie University Code for the Responsible Conduct of Research* and the *Authorship Standard*, researchers have a responsibility to their colleagues and the wider community to treat others fairly and with respect, to give credit where appropriate to those who have contributed to research.

Note for HDR students: *Where research papers are being included in a thesis, this template must be used to document the contribution of authors to each of the proposed or published research papers. The contribution of the candidate must be sufficient to justify inclusion of the paper in the thesis.*

1. DETAILS OF PUBLICATION & CORRESPONDING AUTHOR

Title of Publication (can be a holding title)		**Publication Status** Choose an item.
Further validation of a binaural model predicting speech intelligibility against envelope-modulated noises		☐ In Progress or Unpublished work for thesis submission ☐ Submitted for Publication ☐ Accepted for Publication ■ Published
Name of **corresponding author**	Department/Faculty	**Publication details:** indicate the name of the journal/ conference/ publisher/other outlet
Thibault Vicente	Department of Linguistics - Audiology	Hearing Research

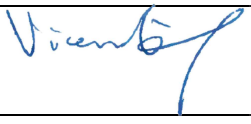
2. STUDENTS DECLARATION (if applicable)

Name of HDR thesis author (If the same as corresponding author - write "as above")	Department/Faculty	Thesis title
as above	Department of Linguistics - Audiology	Modelling the effect of hearing impairment for binaural speech intelligibility in noise

Description of HDR thesis author's contribution to planning, execution, and preparation of the work if there are multiple authors

(for example, how much as a percent did you contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)

This article presents a modelling study including a data collection that was conducted by myself under the supervision of Mathieu Lavandier. The original matlab code of the model was written by Mathieu Lavandier, which was then adapted by myself for the study. Some of the data used for the sensitivity analysis were shared by John Culling and Stephan Ewert. I adapted a matlab code written by a previous student of Mathieu Lavandier to create the stimuli of the experiment. I was the experimenter during the data collection. The data analysis was done by myself after Mathieu Lavandier taught me how to use the statistical software (STATISTICA). The sensitivity analysis was implemented and interpreted by myself. I wrote the first draft of the published paper then I changed it according to the input of Mathieu Lavandier before the submission to Hearing Research. I carried out the publication process with inputs from Mathieu Lavandier.

I declare that the above is an accurate description of my contribution to this publication, and the contributions of other authors are as described below.	**Student signature** **Date**	 03/16/2021

3. Description of all other author contributions

Use an Asterisk * to denote if the author is also a current student or HDR candidate.

The HDR candidate or corresponding author must, for each paper, list all authors and provide details of their role in the publication. Where possible, also provide a percentage estimate of the contribution made by each author.

Name and affiliation of author	**Intellectual contribution(s)** (for example to the: conception of the project, design of methodology/experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content etc.)
Mathieu Lavandier / Univ. Lyon, ENTPE, Laboratoire de Tribologie et Dynamique des	Providing the original matlab code of the model and surpervising the student during the research work. (see student declaration for more details)
	Provide summary for any additional Authors in this cell.

4. Author Declarations

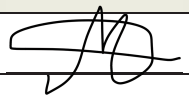
I agree to be named as one of the authors of this work, and confirm:

i. *that I have met the authorship criteria set out in the* Authorship Standard*, accompanying the* Macquarie University Research Code*,*

ii. *that there are no other authors according to these criteria,*

iii. *that the description in Section 3 or 4 of my contribution(s) to this publication is accurate*

iv. *that I have agreed to the planned authorship order following the* Authorship Standard

Name of author	**Authorised *** By Signature or refer to other written record of approval (eg. pdf of a signed agreement or an email record)	**Date**
Mathieu Lavandier		03/16/2021
	Provide other written record of approval for additional authors (eg. pdf of a signed agreement or an email record)	

5. Data storage

The original data for this project are stored in the following location, in accordance with the *Research Data Management Standard* accompanying the *Macquarie University Research Code*.

If the data have been or will be deposited in an online repository, provide the details here with any corresponding DOI.

Data description/format	**Storage Location or DOI**	**Name of custodian** if other than the corresponding author

A copy of this form must be retained by the corresponding author and must accompany the thesis submitted for examination.

MACQUARIE UNIVERSITY
AUTHORSHIP CONTRIBUTION STATEMENT

In accordance with the *Macquarie University Code for the Responsible Conduct of Research* and the *Authorship Standard*, researchers have a responsibility to their colleagues and the wider community to treat others fairly and with respect, to give credit where appropriate to those who have contributed to research.

Note for HDR students: *Where research papers are being included in a thesis, this template must be used to document the contribution of authors to each of the proposed or published research papers. The contribution of the candidate must be sufficient to justify inclusion of the paper in the thesis.*

1. DETAILS OF PUBLICATION & CORRESPONDING AUTHOR

Title of Publication (can be a holding title)		**Publication Status** Choose an item.
A binaural model implementing an internal noise to predict the effect of hearing impairment on speech intelligibility in non-stationary noises		☐ In Progress or Unpublished work for thesis submission ☐ Submitted for Publication ☐ Accepted for Publication ◼ Published
Name of **corresponding author**	Department/Faculty	**Publication details:** indicate the name of the journal/ conference/ publisher/other outlet
Thibault Vicente	Department of Linguistics - Audiology	Journal of the Acoustical Society of America

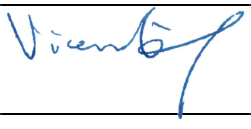
2. STUDENTS DECLARATION (if applicable)

Name of HDR thesis author (If the same as corresponding author - write "as above")	Department/Faculty	Thesis title
As above	Department of Linguistics - Audiology	Modelling the effect of hearing impairment for binaural speech intelligibility in noise

Description of HDR thesis author's contribution to planning, execution, and preparation of the work if there are multiple authors

(for example, how much as a percent did you contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)

This article shows a modelling study that was done by myself under the supervision of Mathieu Lavandier and Jörg Buchholz. The original model was coded by Mathieu Lavandier, but the changes presented in the chapter were done by myself. The internal noise formula presented in the chapter was obtained after long dicussions between the three of us using the outcomes of my literature review on the internal noise. The data used for the chapter were shared by Jörg Buchholz and Baljeet Rana. I wrote the first draft of the published paper and then changed it considering the comments and suggestions of Mathieu Lavandier and Jörg Buchholz before submitting the manuscript to the Journal of the Acoustical Society of America. I carried out the publication process with inputs from Jörg Buchholz and Mathieu Lavandier.

I declare that the above is an accurate description of my contribution to this publication, and the contributions of other authors are as described below.	**Student signature** **Date**	 03/16/2021

3. Description of all other author contributions

Use an Asterisk * to denote if the author is also a current student or HDR candidate.

The HDR candidate or corresponding author must, for each paper, list all authors and provide details of their role in the publication. Where possible, also provide a percentage estimate of the contribution made by each author.

Name and affiliation of author	**Intellectual contribution(s)** (for example to the: conception of the project, design of methodology/experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content etc.)
Mathieu Lavandier / Univ. Lyon, ENTPE, Laboratoire de Tribologie et Dynamique des	Providing the original matlab code of the model and supervising the sudent (see student declaration for more details)
Jörg Buchholz / Department of Linguistics - Audiology, Australian Hearing Hub, Macqu	Providing the data for the study and supervising the sudent (see student declaration for more details)
	Provide summary for any additional Authors in this cell.

4. Author Declarations

I agree to be named as one of the authors of this work, and confirm:

i. *that I have met the authorship criteria set out in the* Authorship Standard*, accompanying the* Macquarie University Research Code*,*

ii. *that there are no other authors according to these criteria,*

iii. *that the description in Section 3 or 4 of my contribution(s) to this publication is accurate*

iv. *that I have agreed to the planned authorship order following the* Authorship Standard

Name of author	**Authorised *** By Signature or refer to other written record of approval (eg. pdf of a signed agreement or an email record)	**Date**
Mathieu Lavandier		03/16/2021
Jörg Buchholz		17/03/2021
	Provide other written record of approval for additional authors (eg. pdf of a signed agreement or an email record)	

5. Data storage

The original data for this project are stored in the following location, in accordance with the *Research Data Management Standard* accompanying the *Macquarie University Research Code*.

If the data have been or will be deposited in an online repository, provide the details here with any corresponding DOI.

Data description/format	**Storage Location or DOI**	**Name of custodian** if other than the corresponding author

A copy of this form must be retained by the corresponding author and must accompany the thesis submitted for examination.

MACQUARIE UNIVERSITY
AUTHORSHIP CONTRIBUTION STATEMENT

In accordance with the *Macquarie University Code for the Responsible Conduct of Research* and the *Authorship Standard*, researchers have a responsibility to their colleagues and the wider community to treat others fairly and with respect, to give credit where appropriate to those who have contributed to research.

Note for HDR students: *Where research papers are being included in a thesis, this template must be used to document the contribution of authors to each of the proposed or published research papers. The contribution of the candidate must be sufficient to justify inclusion of the paper in the thesis.*

1. DETAILS OF PUBLICATION & CORRESPONDING AUTHOR

Title of Publication (can be a holding title)		**Publication Status** Choose an item.
The contribution of binaural unmasking to speech intelligibility in noise for normal-hearing and hearing-impaired listeners		☐ In Progress or Unpublished work for thesis submission ■ Submitted for Publication ☐ Accepted for Publication ☐ Published
Name of **corresponding author**	Department/Faculty	**Publication details:** indicate the name of the journal/ conference/ publisher/other outlet
Thibault Vicente	Department of Linguistics - Audiology	The Journal of the Acoustical Society of America

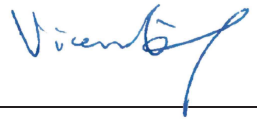
2. STUDENTS DECLARATION (if applicable)

Name of HDR thesis author (If the same as corresponding author - write "as above")	Department/Faculty	Thesis title
As above	Department of Linguistics - Audiology	Modelling the effect of hearing impairment for binaural speech intelligibility in noise

Description of HDR thesis author's contribution to planning, execution, and preparation of the work if there are multiple authors

(for example, how much as a percent did you contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)

This article presents an experimental and modelling study that was designed by myself under the supervision of Jörg Buchholz and Mathieu Lavandier. Regarding the experimental study, I designed the tested conditions with inputs from my supervisors. Kelly Miles and Jörg Buchholz helped me to get the ethic clearance for the data collection. The TFS-AF test (measuring interaural time difference sensitivity) used in the study was developed by Christian Füllgrabe and colleagues. The software to measure intelligibility thresholds was developed by Jörg Buchholz and slighlty changed by myself for our study. The original monaural stimuli were created by Jörg Buchholz in a previous study. The impulse responses were from a previous study by Mathieu Lavandier. I generated the binaural stimuli with this material. I carried out the recruitment of the participants with the help of Kelly Miles, Julie Beadle, Jörg Buchholz and Katrina Freeston. Katrina Freeston was the experimenter because we needed an Australian audiologist to test

I declare that the above is an accurate description of my contribution to this publication, and the contributions of other authors are as described below.	**Student signature** **Date**	 03/16/2021

3. Description of all other author contributions

Use an Asterisk * to denote if the author is also a current student or HDR candidate.

The HDR candidate or corresponding author must, for each paper, list all authors and provide details of their role in the publication. Where possible, also provide a percentage estimate of the contribution made by each author.

Name and affiliation of author	**Intellectual contribution(s)** (for example to the: conception of the project, design of methodology/experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content etc.)
Jörg Buchholz / Department of Linguistics - Audiology, Australian Hearing Hub, Macqu	Helping for the data collection, supervising the student during the experimental and modelling study (see student declaration for more details)
Mathieu Lavandier / Univ. Lyon, ENTPE, Laboratoire de Tribologie et Dynamique des	Supervising the student during the experimental and modelling study (see student declaration for more details)
	Provide summary for any additional Authors in this cell.

4. Author Declarations

I agree to be named as one of the authors of this work, and confirm:

i. *that I have met the authorship criteria set out in the* Authorship Standard*, accompanying the* Macquarie University Research Code*,*

ii. *that there are no other authors according to these criteria,*

iii. *that the description in Section 3 or 4 of my contribution(s) to this publication is accurate*

iv. *that I have agreed to the planned authorship order following the* Authorship Standard

Name of author	**Authorised *** By Signature or refer to other written record of approval (eg. pdf of a signed agreement or an email record)	**Date**
Jörg Buchholz		17/03/2021
Mathieu Lavandier		03/16/2021
	Provide other written record of approval for additional authors (eg. pdf of a signed agreement or an email record)	

5. Data storage

The original data for this project are stored in the following location, in accordance with the *Research Data Management Standard* accompanying the *Macquarie University Research Code*.

If the data have been or will be deposited in an online repository, provide the details here with any corresponding DOI.

Data description/format	**Storage Location or DOI**	**Name of custodian** if other than the corresponding author

A copy of this form must be retained by the corresponding author and must accompany the thesis submitted for examination.

Résumé en français

1 Introduction

La perte auditive touchait 850 millions de personnes en 1990, tandis qu'en 2019, 1,57 milliard de personnes (environ 20 % de la population mondiale) avaient au moins une surdité légère. Cette augmentation est due à la croissance démographique ainsi qu'à l'augmentation de l'espérance de vie car la perte auditive augmente avec l'âge. Cela se traduit par un nombre plus élevé de personnes malentendantes (HI) pour 100 000 personnes (de 15 890 en 1990 à 20 310 en 2019). Le coût d'une perte auditive non traitée est estimé entre 620 et 660 milliards d'euros. Cela comprend les dépenses de santé liées à la perte auditive (à l'exception des coûts des appareils auditifs), la perte de productivité (due au chômage par exemple) et les coûts sociétaux ou intangibles. Une récente revue de littérature a rapporté que souffrir d'une perte auditive à partir de l'âge adulte réduit la qualité de vie et constitue un facteur de risque de détresse en raison de l'isolement social, par exemple. À l'inverse, les personnes qui souffrent de surdité dès le plus jeune âge s'adaptent mieux à leur déficience mais sortent du cadre scolaire prématurément. Les appareils auditifs peuvent en partie réduire l'effet de la perte auditive en restaurant l'intelligibilité et donc la qualité de vie. C'est pourquoi la perte auditive est alors un fardeau sociétal et économique qui va tendre à augmenter au cours des prochaines décennies.

Il est nécessaire dans un premier temps de comprendre quels aspects de la déficience auditive affectent la capacité des personnes HI à interagir avec notre société, pour ensuite fournir des prothèses auditives capable de réduire efficacement l'impact de la perte auditive. En particulier, l'isolement social observé chez les HI suggère qu'ils ont du mal à tenir une conversation dans des environnements bruyants tels que les restaurants, cantines, bars (appelé problème ou situation de «cocktail party»). Des recherches expérimentales ont été menées pour mieux comprendre les mécanismes auditifs qui contribuent à l'intelligibilité de la parole dans les environnements bruyants pour des auditeurs normo-entendants (NH) et HI. Cela a permis de tester, vérifier et concevoir des théories liées à l'intelligibilité de la parole dans le bruit qui ont été plus approfondies par le développement de modèles. Les approches expérimentales et de modélisation se complètent : des données expérimentales sont nécessaires pour concevoir un premier concept de modèle, tandis que ce modèle vérifie le concept s'il décrit les effets observés dans les données, ou aide à réviser ce concept si les prédictions ne décrivent pas l'ensemble des données.

Les résultats de ces recherches suggèrent qu'un auditeur est capable d'isoler le discours cible des sources concurrentes, également appelées masqueurs, lorsqu'ils sont spatialement séparés en utilisant des mécanismes auditifs basés sur les différences interaurales de niveau et de temps (ILD et ITD, respectivement) des signaux. Cependant, le bon fonctionnement de ces mécanismes est altéré lorsqu'un auditeur souffre d'une perte auditive à tel point que l'intelligibilité de la parole peut être fortement réduite. Un certain nombre de modèles issus de la littérature permettent de prédire le bénéfice spatial observé pour les NH lorsque la parole est spatialement séparée des masqueurs (comme dans les situations dites de cocktail party). Peu d'entre eux prennent en compte certains aspects de la perte auditive, d'autres peuvent expliquer l'effet des algorithmes de traitement du signal des prothèses auditives, mais aucun de ces modèles ne fournit de prédictions précises pour un auditeur HI portant des prothèses auditives en environnements réels.

C'est pourquoi la motivation de ce projet est de développer un modèle binaural d'intelligibilité de la parole qui peut prédire des effets liés à la perte auditive dans des scénarios complexes et réalistes, avec de la réverbération et des sources masquantes. Ce modèle aiderait à mieux comprendre quels aspects de la perte auditive et des fonctionnalités des prothèses auditives ont un impact sur l'intelligibilité. L'objectif à long terme est de concevoir de meilleures prothèses (ou de meilleurs algorithmes de traitement du signal) qui permettraient aux HI de communiquer dans des situations bruyantes à l'instar des NH, ce qui améliorerait leurs interactions sociales et réduirait ainsi le fardeau lié à la perte auditive.

La modélisation d'une situation de cocktail party peut être complexe car elle implique différentes sources de masquage, plusieurs profils de perte auditive, plusieurs interlocuteurs, des sources mobiles... Tous ces facteurs ne peuvent pas être abordés dans un même doctorat,

le cadre est donc ici limité aux sources immobiles, avec un seul interlocuteur masqué par des bruits ainsi qu'à certains mécanismes auditifs à la fois pour des NH et pour des HI présentant une perte symétrique, légère à modérément sévère. Les masqueurs contenant de la parole ne sont pas considérés dans le projet malgré leur présence dans les situations de cocktail party car ils impliquent un trop grand nombre de mécanismes perceptifs. La logique est d'utiliser des bruits masquants dans un premier temps pour isoler certains mécanismes perceptifs associés à l'intelligibilité de la parole, puis d'étendre le modèle aux masqueurs contenant de la parole dans un futur projet. La structure de la thèse met en avant le fait que chaque chapitre est une étude à part entière.

2 Etat de l'art

2.1 Compréhension de la parole dans le bruit

2.1.1 Rapport signal sur bruit, intelligibilité et écoute dans les trous

Le rapport signal sur bruit (SNR) est un indice qui permet de comparer le niveau sonore du signal cible (ici, la parole) et d'un bruit. Une augmentation du SNR conduit en général à une augmentation de l'intelligibilité. Les modulations d'amplitude qui se produisent dans les enveloppes des signaux de parole et de masqueurs conduisent à des variations du SNR au cours du temps. Un auditeur peut bénéficier des modulations en amplitude du masqueur en écoutant dans les creux de son enveloppe temporelle, c'est-à-dire lorsque le SNR sur un intervalle de temps est supérieur au SNR moyen.

Une mesure d'intelligibilité qui est très utilisée dans la littérature est le seuil de réception de la parole (SRT). Ce dernier est un SNR qui mène à un certain pourcentage (fixé par défaut dans le manuscrit à 50%) de mots compris par l'auditeur pour une condition acoustique donnée. Il se mesure en fixant le niveau d'un des deux signaux et en variant l'autre pour converger vers le SNR qui mène à 50% de mots compris par l'auditeur.

2.1.2 Démasquage spatial

Le fait d'avoir deux oreilles permet d'avoir accès à des indices binauraux tels que la ILD et ITD. La ILD traduit le fait qu'une source sonore produit un niveau plus élevé à une oreille qu'à l'autre tandis que la ITD quantifie la différence de temps que met le signal de la source à arriver à une oreille comparée à l'autre. L'intelligibilité de la parole dans le bruit peut être améliorée quand la source cible est spatialement séparée des sources bruyantes (comparé à une situation ou les sources seraient co-localisées) grâce à l'utilisation des indices binauraux. Ce démasquage spatial est communément expliqué en utilisant deux mécanismes perceptif: l'écoute à la meilleure oreille et le démasquage binaural.

L'écoute à la meilleure oreille utilise la différence entre la ILD de la cible et la ILD du masqueur pour améliorer l'intelligibilité. Cette différence implique deux SNR aux oreilles de l'auditeur, ainsi il serait capable de profiter du plus grand des deux pour améliorer la compréhension de la parole.

Le démasquage binaural est un mécanisme qui utilise la différence en ITD entre cible et masqueur pour améliorer l'intelligibilité de la parole. Il est expliqué et modélisé dans l'«equalization-cancellation (EC) theory» de Durlach (1972). Ce dernier a supposé qu'un auditeur est capable d'égaliser en interne les signaux masquants à ses oreilles en appliquant un gain et un retard à une oreille pour compenser la ILD et ITD du masqueur. Ensuite, l'auditeur annule partiellement ce masquage en soustrayant les signaux à une oreille à ceux à l'autre oreille pour améliorer le SNR interne et par conséquent l'intelligibilité de la parole. L'efficacité de l'annulation du bruit dépend de la cohérence interaurale du bruit quantifiant la similarité des signaux arrivant aux oreilles de l'auditeur. Plus la cohérence interaurale du bruit est élevée, plus son annulation est efficace.

2.1.3 Effect de la perte auditive

La perte auditive est généralement connue pour induire une élévation des seuils auditifs résultant en une audibilité réduite, ce qui entraîne une difficulté à entendre les sons faibles. La déficience auditive augmente naturellement avec l'âge (presbyacousie) et elle peut être intensifiée à cause

d'une exposition à des niveaux sonores trop élevés. Ces facteurs entrainent des pertes en haute fréquence. L'élévation des seuils auditifs induit la perte d'informations importantes de la parole et des masqueurs, telles que les différences en ILD et ITD entre les sources, dégradant ainsi l'intelligibilité. Une audibilité réduite a un effet direct sur l'avantage procuré par l'écoute à la meilleure oreille, car la parole à l'oreille avec le meilleur SNR peut ne pas être audible ou seulement partiellement audible (Glyde et al., 2013). Une audibilité réduite dégrade probablement la perception de la ITD car lorsqu'un signal n'est pas audible à une oreille, le système auditif ne peut pas conclure sur la ITD du signal.

La déficience auditive ne se limite pas à l'élévation des seuils auditifs, cela implique également d'autres déficits tels que le phénomène de recrutement, la difficulté à détecter un intervalle de temps entre deux signaux, la baisse de la sensibilité à la structure fine des signaux, l'augmentation de la largeur des filtres auditifs... Tous ces déficits tendent à réduire l'avantage du démasquage spatial.

2.2 Modèle de compréhension de la parole dans le bruit

Des modèles de compréhension de la parole dans le bruit ont été développés au cours des dernières décennies. Chaque modèle se différencie par son approche pour prédire l'intelligibilité, le type d'écoute dans lequel il peut être appliqué (monaural ou binaural) et sa capacité à prédire l'effet de la perte auditive. Une description exhaustive de ces modèles est faite en anglais dans le manuscrit (cf. Sec. II.2). Seul les modèles de Collin and Lavandier (2013) et de Lavandier et al. (2018) sont présentés ci-dessous car ce sont les deux modèles utilisés pendant la thèse.

Le modèle de Collin and Lavandier (2013) est conçu pour prédire l'intelligibilité de la parole dans des bruits modulés en amplitude pour des NH. Il prend en entrée le signal cible et les signaux masquants qui sont ensuite décomposés en pas de temps-fréquence. L'écoute à la meilleure oreille est modélisée en prenant le plus grand SNR entre les deux oreilles. L'avantage dû au démasque binaural est calculé en utilisant une formule venant de la littérature qui considère la différence de ITD entre la parole et le bruit ainsi que la cohérence interaurale du bruit. Les valeurs obtenues sont alors intégrées dans le domaine fréquentiel, moyennées dans le domaine temporel et sommées pour obtenir un ratio binaural. Ce dernier peut être converti en un seuil d'intelligibilité pour ensuite être comparé à des mesures de compréhension de la parole.

Lavandier et al. (2018) apporte une modification au modèle précédent pour prendre en compte la perte auditive. Pour cela, l'audiogramme de l'auditeur est désormais une entrée du modèle pour créer un niveau de bruit interne qui est considéré comme un masqueur. Le niveau de bruit interne possède le même spectre que celui de l'audiogramme de l'auditeur et son niveau est fixé par un paramètre du modèle. Dans cette version de modèle, le SNR à la meilleure oreille est calculé en considérant le plus grand niveau entre bruit interne et bruit externe. L'avantage dû au démasquage binaural est calculé seulement si les niveaux des signaux externes sont supérieurs aux niveaux des bruits internes.

2.3 Objectifs de la thèse

Le projet de doctorat s'est concentré sur l'extension des modèles de Collin and Lavandier (2013) et de Lavandier et al. (2018). Le premier modèle a été validé dans la publication originale (Collin and Lavandier, 2013) en utilisant trois jeux de données. Le modèle prédisait avec précision les SRT mesurés cependant, les valeurs des paramètres du modèle n'ont pas été définies rigoureusement et cela devait être pris en compte avant de développer un modèle pour les HI. Le modèle de Lavandier et al. (2018) présente également certaines limites, premièrement, le paramètre définissant le niveau absolu du bruit interne (simulant la perte auditive d'un auditeur) était différent pour les NH et HI afin d'obtenir de bonnes prédictions pour les deux groupes. Cela résulte en deux modèles différents, un pour NH et un pour les HI, ce qui implique que la différence entre les groupes d'auditeurs ne pouvait pas être prédite. De plus, cette version du modèle n'a été validée que sur des ensembles de données impliquant des ILD mais pas de ITD. Par conséquent, la composante du modèle prédisant le démasquage binaural n'a pas encore été testée. Il est important de noter que ces données ont été collectées avec des stimuli anéchoïques, ce qui signifie que le modèle n'a pas encore été testé sur des stimuli plus réalistes incluant de la réverbération d'une salle. Ainsi l'objectif de la thèse est de produire un unique modèle binaural de compréhension de la parole dans des salles bruyantes pour des auditeurs NH et HI.

3 Validation d'un modèle binaural de compréhension de la parole dans des bruits modulés en amplitude pour des auditeurs NH

Le contenu de ce chapitre est publié dans le journal Hearing Research (Vicente and Lavandier, 2020). Collin and Lavandier (2013) ont proposé un modèle binaural pour prédire l'intelligibilité de la parole dans les salles, en présence de sources de bruit modulé en amplitude, pour des NH. Cette version préliminaire possède quatre paramètres dont l'influence n'a pas été testée. Le but de la présente étude a été de réaliser une étude paramétrique du modèle, basée sur une analyse de sensibilité, afin d'optimiser les valeurs des paramètres en utilisant plusieurs expériences de la littérature impliquant des conditions critiques pour tester le modèle.

Conceptuellement, l'analyse de sensibilité consiste à faire des prédictions avec le modèle d'étude en faisant varier la valeur de ses paramètres. Ensuite, les indices de sensibilité peuvent estimer le taux de variance de la sortie du modèle dû à un paramètre donné ou à une interaction entre paramètres. Par exemple, l'indice de sensibilité du premier ordre évalue l'impact direct de la variation d'un paramètre donné sur la sortie du modèle, un indice de sensibilité du second ordre évalue la quantité de variance de la sortie du modèle qui peut être attribuée à une interaction entre deux paramètres, etc. La sortie du modèle considérée ici était soit la corrélation de Pearson (r) ou l'erreur absolue moyenne absolue entre mesures et prédictions ($MeanErr$).

Les quatre paramètres du modèle sont: (1) la durée de la fenêtre de Hann utilisée pour calculer l'avantage lié au démasquage binaural en ms («BU»); (2) la durée de la fenêtre de Hann utilisée pour calculer le SNR à la meilleure oreille en ms («BE», ce sont les deux résolutions temporelles du modèle); (3) le nombre de filtres auditifs utilisé sur une plage fréquentielle donnée («SpecSamp», échantillonage spectral du modèle) exprimé en nombre de filtres par bande passante d'un filtre auditif (ERB); et (4) le paramètre fixant le SNR maximal à la meilleure oreille («Ceiling» en dB). Les valeurs des paramètres étaient fixées à sur 24 ms, 24 ms, 2 filtres par ERB et 20 dB, respectivement (Collin and Lavandier, 2013; Cubick et al., 2018). En particulier, la même résolution temporelle a été utilisée pour modéliser l'écoute à la meilleure oreille et le démasquage binaural; tandis que la résolution temporelle de chaque mécanisme a été étudiée indépendamment ici.

Cinq expériences ont été utilisées pour l'étude basée sur une analyse de sensibilité, dont deux de Culling and Mansell (2013), deux de Collin and Lavandier (2013) et une qui a été réalisée spécifiquement pour l'étude. Culling and Mansell (2013) utilisent des bruits modulés artificiellement par un signal carré, avec différentes fréquences de modulations, tout en isolant les composantes du démasquage spatial, ainsi il était possible de tester la résolution temporelle de chacune d'elles. Les expériences de Collin and Lavandier (2013) utilisent des bruits modulés par des enveloppes de signaux de parole. La première était pertinente pour tester la prédiction du bouchage des trous par la réverbération, le fait que la queue de réverbération va niveler les modulations des signaux. La deuxième expérience permettait d'étudier la prédiction du démasquage spatial à réverbération constante. L'expérience qui a été réalisée pendant ce travail a permis de tester simultanément l'influence de la réverbération sur le bouchage des trous, en isolant les deux composantes binaurales du modèle, avec des masqueurs présentant des modulations semblables aux signaux de la parole.

Les conclusions de l'analyse de sensibilité étaient similaires lorsque r ou $MeanErr$ était considérée comme sortie du modèle. Les valeurs des indices de sensibilité étaient différentes mais les tendances observées étaient les mêmes. Le paramètre avec le plus d'influence directe était Ceiling pour quatre expériences et BE pour la cinquième. Le seul indice de sensibilité du second ordre non négligeable était celui lié à l'interaction entre Ceiling et BE. Dans toutes les expériences la variation de $MeanErr$ était presque entièrement due aux impacts directs des paramètres et à l'interaction entre BE et Ceiling.

Aucune raison évidente dans les prédictions laissait penser qu'il fallait modifier la valeur de SpecSamp donc sa valeur initiale a été gardée. Concernant BU, augmenter sa valeur de 24 ms à 300 ms permettait de mieux prédire les données Culling and Mansell (2013) sans dégrader les prédictions des autres expériences. La valeur de Ceiling et celle de BE ont été choisies en même temps à cause de l'interaction mesurée entre ces paramètres. Le choix final a été de ne pas modifier les valeurs initiales de ces paramètres puisqu'elles faisaient partie des valeurs menant à des prédictions précises. Une fois le modèle révisé, il a été testé sur un nouveau jeu de données (Ewert et al., 2017) afin d'évaluer ses capacités prédictives. Ce jeu de données étudie l'influence

spectro-temporelle des masqueurs sur le démasquage spatial en considérant 40 conditions. Le modèle révisé prédisait avec précision les données des 6 expériences, avec r variant entre 0,85 et 0,96; et $MeanErr$ variant entre 0.5 et 1.4 dB, selon l'expérience considérée. De plus, les valeurs des paramètres du modèle ont été minutieusement testées et choisies.

4 Un modèle binaural implémentant un bruit interne pour prédire l'effet de la perte auditive sur la compréhension de la parole dans des bruits modulés en amplitude

Le contenu de ce chapitre est publié dans le Journal of the Acoustical Society of America (Vicente et al., 2020). Dans cette étude, le modèle développé par Lavandier et al. (2018) a été modifié afin qu'un seul modèle soit utilisé pour prédire l'intelligibilité de la parole pour des NH et HI au lieu de modèles séparés (comme expliqué dans la section 2.3).

Le changement entre le modèle proposé et le modèle de Lavandier et al. (2018) a porté seulement sur une nouvelle implémentation du bruit interne (bruit visant à modéliser la perte auditive d'un auditeur). Il se compose désormais de deux composantes: l'une liée aux seuils auditifs et l'autre liée à l'effet du niveau sonore externe sur l'intelligibilité. Plus précisément, le niveau de bruit interne augmente avec le niveau sonore externe. Trois hypothèses ont été formulées pour cette nouvelle implémentation. (1) Le niveau global des stimuli externes est approximé par le niveau du masqueur connu (car fixe dans la mesure des SRT), en supposant qu'ainsi le SNR large bande est inférieur à 0 dB. (2) La perte auditive de l'auditeur peut être divisée en deux proportions (η et $1-\eta$) pour refléter les différentes contributions des pertes de cellules ciliées externes et internes (OHC et IHC, respectivement); η étant identique à toutes les fréquences et pour tous les auditeurs. (3) La valeur maximale autorisée pour la perte estimée des OHC est de 57,6 dB (Moore and Glasberg, 2004) et la perte au-dessus de cette valeur contribue uniquement à la perte estimée des IHC. La perte des OHC sert à créer un seuil minimal de bruit interne car ces cellules sont liées à l'audibilité des signaux. La perte des IHC est utilisée pour créer la composante du bruit interne qui est liée au bruit externe car ces cellules sont liées à la façon dont les signaux sont codés. En plus de η, deux autres paramètres sont implémentés dans la formule du bruit interne qui servent à définir le niveau du seuil minimal de bruit interne et le niveau de bruit externe à partir duquel le niveau de bruit interne commence à augmenter. Les valeurs de ces trois paramètres ont été variées puis fixées pour obtenir les meilleures prédictions sur trois jeux de données de la littérature.

Cette implémentation s'est inspirée des conclusions de Bernstein and Trahiotis (2008), qui fournissent une revue de la littérature sur le concept de bruit interne avant de mener des expériences pour caractériser le bruit interne chez des NH. Conformément à leur revue de la littérature, leurs résultats suggèrent que le bruit interne serait constitué de deux composantes. La première composante est indépendante des stimuli et détermine le seuil auditif absolu (similaire au seuil minimal de bruit interne dans l'implémentation proposée). La seconde composante dépend des stimuli et voit son niveau augmenter lorsque le niveau de bruit externe augmente à partir d'un certain niveau et en suivant une règle d'un dB pour un dB.

Le modèle a été validé sur trois expériences mesurant des SRT avec des NH et HI. Les stimuli étaient anéchoïques et joués au casque. La source cible, simulée devant l'auditeur, a été présentée simultanément avec deux bruits masquants possédant le même spectre moyen que celui de la cible. Un masqueur n'était pas modulé en amplitude tandis que l'autre possédait des modulations qui s'apparentaient à des modulations de deux discours superposés. Deux séparations cible-masqueur étaient considérées, soit les masqueurs étaient co-localisés avec la cible, soit ils étaient séparés à $\pm$ 90°. Plusieurs niveaux de bruit ont été testés ainsi que deux types d'amplification linéaire. De plus, deux expériences mesurant des SRT seulement avec des NH ont été impliquées dans l'étude pour vérifier la rétrocompatibilité du modèle avec ses versions précédentes (Lavandier et al., 2012; Collin and Lavandier, 2013). Les prédictions du modèle (une fois les valeurs des paramètres du bruit interne définis) sont précises sur les cinq expériences menant à une corrélation r supérieure ou égal à 0,93; et $MeanErr$ ne dépassant pas 1,1 dB. Même si le modèle a prédit les données avec succès, sa composante liée au démasquage binaural n'a été testée que dans une seule condition acoustique sur les trois expériences impliquant des HI. Cela s'avère insuffisant pour prédire complètement le démasquage binaural chez les HI. C'est pourquoi le prochain chapitre s'attèle entre autre à la validation de la composante du modèle

liée au démasquage binaural.

5 La contribution du démasquage binaural pour l'intelligibilité de la parole dans le bruit pour des NH et HI

Le contenu de ce chapitre est soumis au Journal of the Acoustical Society of America. Ce chapitre vise à étudier rigoureusement les effets liés à la perte auditive, de la réverbération (qui fait varier la cohérence interaurale des signaux), du niveau sonore, de l'enveloppe du masqueur et de la différence d'azimut cible-masqueur sur la contribution du démasquage binaural pour l'intelligibilité de la parole dans le bruit. Pour cela des SRT ont été mesurés (en faisant varier ces précédents facteurs) ainsi que la sensibilité des auditeurs à détecter une ITD (en utilisant le test de Füllgrabe and Moore, 2017) afin d'étudier la relation entre cette sensibilité et le démasquage binaural.

Douze NH et huit HI ont été recrutés pour l'étude expérimentale. Les SRT ont été mesurés sous 22 conditions acoustiques. Trois niveaux de réverbération ont été testés et appliqués seulement sur le bruit pour faire varier sa cohérence interaurale. La cible n'a pas été pas soumise à la réverbération car cela implique des effets acoustiques qui n'auraient pas été pertinents pour l'étude. Les deux mêmes bruits que dans le chapitre précédent ont été considérés. Un seul masqueur était présenté simultanément avec la cible. Le niveau du masqueur (fixe pendant la mesure d'un SRT) pouvait prendre une des trois valeurs testées dans l'expérience. Trois séparations cible-masqueur ont été testés en fixant la position du masqueur et variant l'azimut de la cible. Les signaux ont été joués au casque afin de simuler la salle virtuelle. Une amplification linéaire a été appliquée aux stimuli joués aux HI, qui considérait leurs pertes auditives par bande de fréquences pour définir un gain. La tête de l'auditeur a été simulée avec deux microphones omnidirectionnels afin de minimiser les effets de l'écoute à la meilleure oreille (car les ILD induit par la tête étaient supprimées) et ainsi d'isoler les effets liés au démasquage binaural.

L'analyse statistique conduite sur les SRT a montré que tous les facteurs testés avaient un effet significatif sur l'intelligibilité. La diminution du coefficient d'absorption de la salle (augmentation de la réverbération) a entrainé une diminution de la cohérence interaurale du masqueur et ainsi une diminution significative de l'intelligibilité. L'augmentation de la séparation masqueur-cible (c'est-à-dire augmentation de la différence de ITD) a augmenté significativement l'intelligibilité, ce qui est expliqué par la EC theory de Durlach (1972). La diminution du niveau sonore a diminué significativement l'intelligibilité (effet de l'audibilité). Le masqueur modulé en amplitude masquait significativement moins la parole que le masqueur stationnaire en amplitude. Cela signifie que les auditeurs ont eu recours à l'écoute dans les trous.

Une autre analyse statistique cette fois menée sur l'avantage dû au démasquage binaural a montré que les HI présentaient en moyenne significativement moins d'avantage que les NH. Une corrélation significative a été reportée entre l'avantage moyen dû au démasquage binaural et la sensibilité des auditeurs à une ITD en considérant les auditeurs des deux groupes. Cette corrélation n'était plus significative au sein de chaque groupe d'auditeurs, ce qui signifie que cette mesure était assez sensible pour expliquer la différence d'intelligibilité entre groupes mais pas entre auditeurs au sein d'un même groupe.

Le modèle développé dans le chapitre précédent prédisait avec précision l'effet de la séparation masqueur-cible, de la réverbération sur le bruit et du type de masqueurs sur le démasquage binaural ($r = 0,92$; $MeanErr = 1,1$ dB). Cependant l'effet du niveau sonore n'était pas bien prédit, en particulier pour les NH. Ainsi, une modification de la formule qui est utilisée pour calculer l'avantage lié au démasquage binaural a été proposée. Cette formule contient deux paramètres qui reflètent l'erreur que fait le système auditif pendant l'égalisation des signaux dans l'EC theory. La modification consista à rendre ces deux paramètres dépendant du niveau sonore externe de sorte qu'ils augmentent quand le niveau diminue, ainsi traduisant le fait que le système auditif est moins précis dans l'égalisation des signaux quand leur audibilité diminue. Cela a permis d'améliorer la prédiction de l'effet du niveau sonore sur le démasquage binaural.

Une discussion est faite dans le chapitre sur le fait que le bruit interne proposé n'a pas de composante binaurale et est défini indépendamment à chaque oreille. Cependant il permet tout de même de prédire l'avantage lié au démasquage binaural. Cela est différent des autres modèles de la littérature (par exemple, Bernstein and Trahiotis, 2017; Beutelmann et al., 2010) qui ont implémenté une composante binaurale à leurs bruits internes.

6 Conclusion

Les connaissances sur le système auditif ont été étendues au cours du projet grâce à des études expérimentales et de modélisation. Elles ont permis de mieux comprendre les mécanismes de l'écoute à la meilleure oreille et du démasquage binaural et leurs relations avec la perte auditive. La dernière version de modèle proposée (développée dans le chapitre V) est capable de prédire l'avantage dû à l'écoute à la meilleure oreille et au démasquage binaural pour les NH et HI. Ceci est principalement réalisé en modélisant l'audition de l'auditeur par un niveau de bruit interne basé sur l'audiogramme. Il est défini indépendamment à chaque oreille (c'est-à-dire sans tenir compte d'une quelconque caractéristique binaurale) et possède deux composantes qui prennent en compte l'influence de la perte des OHC et des IHC ainsi que l'effet du niveau des stimuli externes.

Le chapitre III a permis de tester avec rigueur l'influence des paramètres du modèle ainsi que de leur fixer une valeur afin de réduire les erreurs entre données prédictions. Cela a également montré que la résolution temporelle pour calculer le démasquage binaural devait être plus grande que celle pour calculer le SNR à la meilleure oreille afin d'obtenir de meilleures prédictions.

Le chapitre IV a proposé une modification du modèle développé par Lavandier et al. (2018), qui permet de prédire l'intelligibilité pour les auditeurs NH et HI à divers niveaux sonores en utilisant une seule version de modèle. Ceci a été accompli grâce au développement d'un niveau de bruit interne à chaque oreille de l'auditeur pour simuler son audition. Le concept de bruit interne était déjà connu dans la littérature mais une implémentation telle que proposée dans l'étude est une nouveauté. Elle considère une séparation de l'audiogramme en perte de OHC et IHC pour créer un seuil minimal de bruit interne (basé sur la perte des OHC) ainsi qu'un bruit interne additif (basé sur la perte des IHC) qui augmente lorsque le niveau des stimuli externe augmente. Ainsi l'étude actuelle a considérablement élargi le champ d'application de la définition d'un bruit interne en deux composantes. Cela signifie que la perte auditive peut être considérée comme un bruit dépendant du niveau sonore externe, qui se trouve dans le système auditive, limitant ainsi les capacités à analyser et à séparer la parole du bruit.

Le chapitre V a étudié la contribution du démasquage binaural pour l'intelligibilité de la parole pour des NH et HI. Les données expérimentales collectées spécialement pour l'étude ont démontré que les auditeurs HI présentaient un avantage dû au démasquage binaural réduit par rapport aux auditeurs NH. La diminution de ce même avantage lorsque le niveau sonore diminue a été confirmée par notre étude. Les résultats de la modélisation ont également été importants. Premièrement, la formule développée par Durlach (1972, qui a été utilisée par Culling et al., 2004, 2005) a été modifiée pour tenir compte de l'effet du niveau sonore des stimuli. Pour ce faire, les paramètres de la formule ont été modifiés pour qu'ils varient en fonction du niveau de sonore externe. Cela met en évidence le fait que le système auditif fait probablement plus d'erreurs dans l'égalisation des signaux (cf. EC theory) lorsque l'audibilité diminue. Deuxièmement, appliquer la formule modifiée uniquement lorsque les signaux sont audibles dans une bande de fréquences (c'est-à-dire lorsque le niveau de chaque signal est au-dessus du niveau du bruit interne à chaque oreille) a permis de prédire le démasquage binaural pour les auditeurs NH et HI. Troisièmement, considérer un bruit interne sans composante binaurale est suffisant pour prédire le démasquage binaural avec le modèle.

La dernière version de modèle doit être maintenant étendue à plus de HI en considérant plus de profils auditifs et l'influence des prothèses auditives. Pour le moment, le modèle a été utilisé seulement pour prédire la différence d'intelligibilité en moyenne sur les auditeurs de chaque groupe. Il faudra donc par la suite considérer la prédiction d'intelligibilité pour un HI seulement.