A binaural model predicting the effect of hearing impairment and noise level on speech intelligibility

Thibault VICENTE$^{(1,2)*}$, Mathieu LAVANDIER$^{(1)}$, Jörg M. BUCHHOLZ$^{(2)}$

$^{(1)}$Univ Lyon, ENTPE, Laboratoire Génie Civil et Bâtiment, Rue Maurice Audin, 69518 Vaulx-en-Velin Cedex, France
$^{(2)}$Department of Linguistics - Audiology, Australian Hearing Hub, Macquarie University, 2109 NSW, Australia

Abstract

A model will be presented that allows predicting binaural speech intelligibility in noise for normal-hearing (NH) and hearing-impaired (HI) listeners. The model inputs are the masker and target signals at the listener's ears, as well as their audiogram that is used to implement an internal noise (IN). The model computes per time frame and frequency band a binaural masking level difference (BMLD) when all signal levels are higher than the IN level, and the signal-to-noise ratio at the better ear (BE-SNR) using the maximum between the masker and IN levels. The BMLD and BE-SNR are integrated across time and frequency and summed to obtain a binaural ratio. The relative differences of the binaural ratios are compared to speech reception threshold (SRT) differences measured in listening tests. The model was validated on one experiment involving NH and HI listeners. Stimuli were anechoic and played using headphones. The target, in front of the listener, was presented simultaneously with two noise-vocoded speech maskers either collocated with the target or separated at +/- 90°. The noise was presented at different sensation levels and the relative target level was adapted to derive SRTs. The experimental data were accurately predicted by the model.

Keywords: Model, Speech Intelligibility, Binaural, Hearing Impairment

1 INTRODUCTION

Having two ears is helpful to improve target intelligibility in noisy environment. A perceptual mechanism well-know as "Spatial Release from Masking" (SRM) quantifies our ability to benefit from binaural cues, the interaural level differences (ILD) and interaural time differences (ITD) (e.g. [7]), to improve target intelligibility when it is spatially separated from the masker. However, SRM is reduced for HI listeners [8].

In order to predict the influence of the noise on speech intelligibility, several models have been developed for NH listeners [16, 5] and for HI listeners [4, 10]. The study will focus on the model developed by Lavandier *et al.* [10], which allowed to predict binaural speech intelligibility in the presence of envelope modulated noises for NH and HI listeners. However, they had to use two different model parameter values for each of the two listener groups to obtain good predictions, effectively resulting actually in two different models. The current study proposes a modification to the concept of the IN (i.e, the model component that describes the listener's hearing ability) allowing to predict data involving NH and HI listeners [14] with a single model.

2 MODEL DESCRIPTION

An updated version of the model developed by Lavandier *et al.* [10] is presented in this study which allows predicting the effect of hearing impairment on binaural speech intelligibility for NH and HI listeners. In brief, masker and target signals at the listener's ears are taken as inputs for the model as well as the listener's audiogram in order to build an IN which is considered as another masker. Based on the incoming signals and the IN, the binaural unmasking advantage and the BE-SNR are computed per time frame and frequency band,

---

[*] Corresponding address: thibault.vicente@entpe.fr

averaged across time and frequency, and added to obtain a binaural ratio. The differences between binaural ratios are then compared to the differences between the SRTs measured in the listening test.

The target energy is averaged across time to avoid that a lack of energy leads to a low signal-to-noise ratio (SNR) and thus, a poor intelligibility, even though it could correspond to a pause between two words, which provides relevant information for the listener. The masker signal energy is taken into account as a function of time to allow the model to predict the ability of a listener to understand speech in the dips of the masker's envelope. The long-term characteristics (i.e., spectra at each ear and interaural phase differences) of the target are computed once, and then combined with those of the masker resulting from the time-frame analysis. Before all signals are passed through a gammatone filterbank with two filters per equivalent rectangular bandwidth, the masker signals are segmented using a 24-ms and also a 300-ms half-overlapping Hann window [15].

The IN is implemented based on the listener's audiogram. The binaural unmasking advantages are computed by applying the BMLD formula from [6] on the signals from the longer time window only if the masker and target levels are above the IN level. On the other hand, the BE-SNR is computed choosing the higher SNR between the left and right ear (with a ceiling SNR value of 20 dB to avoid that the SNR goes to infinity in the masker's dips) while considering the higher level between the IN and the masker signal using the shorter window. Then, the model averages across time and frequency (using a SII weighting, [1]) the binaural unmasking advantages and BE-SNRs and adds these values to obtain the binaural ratio.

In order to compare the target and masker signals and the listener's IN, the signals are set at the actual levels played during the experiment, in dB SPL. As for the design of the IN spectrum at each ear, three assumptions are made, (1) the overall level of the external stimuli is approximated by the masker level, in other words it is assumed that the broadband SNR is below 0 dB, (2) the hearing loss (HL) is split into X % of outer hair cell loss (HL_{OHC}) and 1-X % of inner hair cell loss (HL_{IHC}) at each frequency and (3) the maximum allowed for HL_{OHC} is 57.6 dB [11], the loss above this value contributes to the HL_{IHC}. The HL_{OHC} and HL_{IHC} are interpolated at the model's center frequencies between the lower and upper center frequency of the audiogram and extrapolated otherwise, using a logarithmic frequency scale. The IN is then calculated using the following formula:

$$InternalNoise(n,N) = aud_X(n) + 10 * log10(10^{\frac{B}{10}} + 10^{\frac{N-Nlim+(1-X)HL(n)}{10}}) \tag{1}$$

where, n is the center frequency of the n^{th} frequency band and N is the overall (external) masker level (averaged across listerner's ears). The term aud_X is defined by $aud_X(n) = HL2SPL(n) + X * HL(n) = HL2SPL(n) + HL_{OHC}(n)$ in dB SPL, with HL2SPL the transformation to convert hearing loss from dB HL into dB SPL. This transformation results from the sum of the reference equivalent sound pressure levels for the THD 39 headphones applied in measuring the subject's audiograms [9] and nominal values for the transformation from 6 cc coupler to ear drum levels [3]. The transformation is interpolated between 200 Hz and 6 300 Hz (the range where values are available) and extrapolated otherwise to get the transformation values at the center frequencies of the model. The term $(1-X)HL(n)$ represents the HL_{IHC}, in dB HL. The free parameters B, Nlim, X were tested before being set at -10 dB, 83 dB and 70%, respectively (see section 4).

For the predictions of the data presented below, the target signal was created by averaging the waveforms of 128 sentences used during the listening test. All sentences were thereby truncated to the shortest sentence duration. The duration of each masker signal was 2 minutes. All signals were convolved with the impulse response of the (equalized) headphones used for data collection and measured on a 4128C Bruel&Kjaer head and torso simulator. Then all signals were calibrated to the sound levels used in the experiment.

3 DATA

To validate the model, an experiment was used that specially investigated the effect of audibility on SRT [14]. Ten NH listeners (mean age of 23.2 +/- 3.2 years, hearing loss <15 dB HL) participated in the experiment as well as ten sensorineural HI listeners (mean age of 70.3 +/- 7.8 years). They had a symmetric (threshold difference between ears <10 dB) mild to moderate, sloping sensorineural hearing loss. Their four-frequency

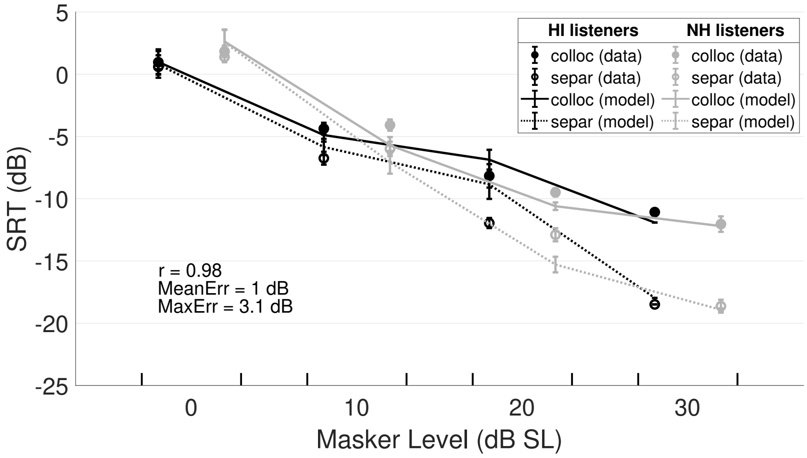

Figure 1. Mean SRTs with standard errors across the NH listeners (in grey symbols) and HI listeners (in black symbols) measured and predicted for four overall masker sensation levels. The two maskers were two noise-vocoded speech either placed at the same position as the target in front of the listener (colloc) or on each side of the listener at +/- 90° (separ).

(0.5, 1, 2, 4 kHz) average hearing loss (4FAHL) was 29.1 +/- 8.0 dBHL. All participants were English native speakers.

The SRTs were measured using BKB-like sentences [2] and an adaptative procedure. The target was simulated in front of the listener. Two noise-vocoded speech maskers were simultaneously presented with the target either collocated with it or symmetrically placed on each side of the listener at +/- 90°. The frontal position was simulated by convolving the anechoic stimuli with an head-related transfer function for frontal incidence averaged across ears, resulting in diotic listening. The spatially separated configuration was simulated by presenting the left masker only through the left headphone and the right masker through the right headphone. Target and maskers signals were filtered, considering nine frequency regions, in order to equalize individual audibility across listeners and they were played at four different sensation levels (0, 10, 20, 30 dB SL, reference: individual SRT in quiet). However, some HI listeners could not be tested for some of the higher sensation levels due to loudness discomfort. All stimuli were presented through the Sennheiser HD215 circumaural headphones and were filtered in order to have the same long-term spectrum as the averaged spectrum of the 1280 sentences from the corpus after applying the above diotic spatialization process.

Figure 1 shows measured SRTs plotted as a function of the masker level in dB SL. The grey and black circles represent the measured SRTs for the NH and HI listeners, respectively. The collocated configuration (colloc) is displayed as filled circles and as open circles for the separated configuration (separ). Increasing the overall sensation level always led to lower SRTs (i.e., better intelligibility) as well as an increase in SRM as measured by the difference in SRTs between the collocated and spatially separated conditions. No significant SRT differences between the two groups of listeners were found, suggesting that performance was similar across groups when audibility was controlled carefully.

4 MODEL PREDICTIONS

The model predictions are plotted in figure 1 using solid and dashed lines corresponding to the spatially collocated and separated conditions, respectively, either in grey for the NH listeners or in black for the HI listeners.

First, in order to find the best combination of parameters in formula 1 (i.e. B, Nlim and X), three indices of model performance were used: the Pearson's correlation coefficient r between data and prediction, the mean absolute error (MeanErr, computed as the average across conditions of the absolute difference between data and prediction) and the maximum absolute error (MaxErr, computed as the maximum of the absolute difference between data and predictions across all conditions). The free parameters B, Nlim and X were varied within ranges [-16;-8] dB, [65;85] dB and [60;90]% [12], respectively in order to simultaneously minimize the two types of error and maximize the correlation. The best predictions were obtained with the values of B, Nlim and X equal to -10 dB, 83 dB and 70 %, respectively. Using these values, the model predicts well the data with a correlation, mean absolute error and maximum absolute error equal to 0.98, 1 dB and 3.1 dB, respectively.

5 DISCUSSION

The revised model allowed to accurately predict the effect of audibility on speech intelligibility for NH and HI listeners, using a single model. This is in contrast to the previous version of the model, developed by Lavandier *et al.* [10], which required a different parameter value for the IN to predict the effect of audibility on speech intelligibility for NH listeners as well as for HI listeners, which effectively resulted in two different models. The new implementation of the IN depends on the external noise level and takes into account the different effects of inner and outer hair cell loss on speech intelligibility. The resulting model provided intelligibility predictions that were as accurate as the ones provided by the two previous models [10].

Considering these encouraging results, the revised model needs to be further validated using additional intelligibility data. For instance, the data presented by [13] could be used to evaluate the performance of the model to predict the difference in both SRT and SRM between stationary speech-shaped noise and (modulated) noise-vocoded speech maskers for NH as well as HI listeners. Given that the stimuli that were used so far for evaluating the proposed model did not involve realistic ITDs, future research will need to specifically evaluate the model-internal processes that are related to ITDs. In particular, this should include an analysis of the potential effect of reduced ITD sensitivity in HI listeners on binaural unmasking.

ACKNOWLEDGEMENTS

The PhD of Thibault Vicente is funded by la Fondation pour l'Audition and Macquarie University (through an international research excellence scholarship, iMQRES). The participation of Thibault Vicente at ICA 2019 was funded by la Fondation pour l'Audition, la Société Française d'Acoustique and by the ICA-ASA-DEGA Young Scientist Conference Attendance Grants (YS Grant). This work was performed within the Labex CeLyA (ANR-10-LABX-0060/ANR-16-IDEX-0005).

REFERENCES

[1] ANSI, American Standard ASA S3.5 : Methods for calculation of the speech intelligibility index, New-York, 1997

[2] Bench J.; Kowal A.; Bamford J. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. British Journal of Audiology, Vol 13 (3), 1979, pp 108–12.

[3] Bentler R. A.; Pavlovic C. V. Transfer Functions and Correction Factors Used in Hearing Aid Evaluation and Research Free Field to Eardrum Transfer Function, Ear and Hearing, Vol 10 (1), 1989, pp 58–63.

[4] Beutelmann R.; Brand T.; Kollmeier B. Revision, extension, and evaluation of a binaural speech intelligibility model, Journal of the Acoustical Society of America, Vol 127 (4), 2010, pp 2479–2497.

[5] Collin B.; Lavandier M. Binaural speech intelligibility in rooms with variations in spatial location of sources

and modulation depth of noise interferers, Journal of the Acoustical Society of America, Vol 134 (2), 2013, pp 1146–1159.

[6] Culling J. F.; Hawley M. L.; Litovsky R. Y. Erratum: The role head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources, Journal of the Acoustical Society of America, Vol 118 (1), 2005, pp 552.

[7] Ewert S.D.; Schubotz W.; Brand T.; Kollmeier B. Binaural masking release in symmetric listening conditions with spectro-temporally modulated maskers, Journal of the Acoustical Society of America, Vol 142 (1), 2017, pp 12–28.

[8] Glyde, H.; Cameron, S.; Dillon, H.; Hickson, L.; Seeto M. The effects of hearing impairment and aging on spatial processing, Ear and Hearing, Vol 34 (1), 2013, pp 15–28.

[9] ISO, International Standard 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, Geneva, 1994.

[10] Lavandier, M.; Buchholz, J. M.; Rana, B. 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, Vol 104 (5), 2018, pp 909-913.

[11] Moore B. C. J.; Glasberg B. R. A revised model of loudness perception applied to cochlear hearing loss, Hearing Research, Vol 118, 2004, pp 70–88.

[12] Pieper I.; Mauermann M.; Oetting D.; Kollmeier B.; Ewert S. D. Physiologically motivated individual loudness model for normal hearing and hearing impaired listeners, Journal of the Acoustical Society of America, Vol 144 (2), 2018, pp 917–930.

[13] Rana B.; Buchholz J. M. Better-ear glimpsing at low frequencies in normal-hearing and hearing-impaired listeners, Journal of the Acoustical Society of America, Vol 140 (2), 2016, pp 1192–1205.

[14] Rana B.; Buchholz J. M. Effect of audibility on better-ear glimpsing as a function of frequency in normal-hearing and hearing-impaired listeners, Journal of the Acoustical Society of America, Vol 143 (4), 2018, pp 2195–2206.

[15] Vicente T.; Lavandier M.. Further validation of a binaural model predicting speech intelligibility against envelope-modulated noises, Hearing Research, under review.

[16] Wan R.; Durlach N. I.; Colburn, H. S. Application of a short-time version of the Equalization-Cancellation model to speech intelligibility, Journal of the Acoustical Society of America, Vol 136 (2), 2014, pp 768-776.