Acta Acustica united with Acustica
Vol. 104 (2018) 909 – 913

DOI 10.3813/AAA.919243

A Binaural Model Predicting Speech Intelligibility in the Presence of Stationary Noise and Noise-Vocoded Speech Interferers for Normal-Hearing and Hearing-Impaired Listeners

Mathieu Lavandier[1)], Jörg M. Buchholz[2)], Baljeet Rana[2)]

[1)] Univ Lyon, ENTPE, Laboratoire Génie Civil et Bâtiment, Rue M. Audin, F-69518 Vaulx-en-Velin, France. mathieu.lavandier@entpe.fr

[2)] Department of Linguistics - Audiology, Australian Hearing Hub, Macquarie University, Sydney, Australia

Summary

A binaural model is presented which predicts the effect of audibility on the intelligibility of speech in the presence of speech-shaped noise and vocoded-speech maskers. It takes the calibrated target and masker signals (independently) and the listener's tonal audiogram at each ear as inputs. Model predictions are compared to speech reception thresholds (SRTs) measured for normal-hearing (NH) and hearing-impaired (HI) listeners in the presence of two uncorrelated speech-spectrum noises or two vocoded-speech maskers, which were either (artificially) spatially separated or co-located with the frontal speech target. The artificial spatial separation was realized by presenting each masker to a different single ear using headphones, while the target was presented diotically as coming from the front. Audibility was varied by testing four different sensation levels for the combined maskers. The model allows for a good prediction of the decrease of SRT and the increase of spatial release from masking (based primarily on better-ear glimpsing here) with increasing audibility. For both groups of listeners, the averaged absolute prediction error across conditions was between 0.6 and 1.7 dB.

PACS no. 43.66.Ba, 43.66.Pn, 43.66.Sr, 43.71.An, 43.71.Ky

1. Introduction

Possessing two ears is useful for understanding speech in noise. It allows for a competing sound source to cause less masking when it is spatially separated from the target. This spatial release from masking (SRM) relies on interaural level and time differences in the signals reaching the ears (ILDs and ITDs, respectively) [1, 2]. It has been shown that SRM is substantially reduced for HI listeners [3]. While several binaural models exist to describe SRM for NH listeners (see [2] for references), we are aware of only one model proposed to predict SRM also for HI listeners. This model has been tested using SRTs measured in the presence of a single noise masker, which was either stationary [1] or modulated in amplitude [4]. The aim of the present study was to propose an alternative modeling approach to predict the effect of audibility on SRTs and SRM for both NH and HI listeners. The performance of this model was tested here for speech presented against two maskers (producing both stationary noise or vocoded speech).

Received 27 February 2018,
accepted 10 July 2018.

2. Model

The proposed model is based on an updated implementation of the model of Collin and Lavandier [2], which predicts binaural speech intelligibility in the presence of multiple non-stationary noises, but does not take hearing impairment and audibility into account. It combines the effects of better-ear listening and binaural unmasking and is based on two inputs: the ear signals generated by the target and those generated by the sum of all interferers. Based on these inputs, the model computes the better-ear signal-to-noise ratio (SNR), as the maximum value of the SNR at the left and right ears, and the binaural unmasking advantage (in dB) from the target and masker interaural parameters. The computation is realized in frequency bands and followed by integration across bands. Adding the better-ear and binaural unmasking components, the model finally produces a (broadband) "effective binaural ratio". Binaural ratios are inverted in order to be compared to SRTs, so

that high ratios correspond to low thresholds (high intelligibility).

The predictions are based on short-term predictions averaged across time. To avoid target speech pauses mistakenly leading to a reduction in predicted intelligibility, the model considers interfering energy as a function of time and target energy averaged across time. Instead of replacing the target speech by a stationary signal with similar long-term spectrum and interaural parameters and applying the short-term analysis on this signal [2], the present implementation computes the long-term statistics of the target only once and combines those with the short-term statistics of the noise to compute binaural ratios within each time frame (before averaging). The model uses 24-ms half-overlapping Hann windows as time frames [2] and a gammatone filterbank with center frequencies ranging from 30 to 19885 Hz and two filters per equivalent rectangular bandwidth (ERB). A ceiling corresponding to the maximum better-ear SNR allowed by frequency band and time frame is also applied to avoid this SNR tending to infinity in interferer pauses. The ceiling value was set to 20 dB[1].

In order to take into account the effects of audibility for both NH and HI listeners, the model proposed here introduces several modifications to the initial model of Collin and Lavandier. The absolute broadband level of the target and masker signals and the listener's tonal audiogram, all expressed in dB sound pressure level (SPL), are required as additional inputs for the model. Predictions are computed separately for each listener (results below are averaged across listeners). While computing the binaural ratios, the target and masker levels are compared to the levels of internal noises which are based on the listener's hearing thresholds. For the better-ear component evaluation, the SNR is computed in each time frame, frequency band and at each ear by subtracting from the target level the maximum between the masker and internal noise levels. The binaural unmasking advantage is set to 0 dB as soon as the masker or target level is below the internal noise level at one ear; otherwise this component of the model is not modified. As discussed below, this rather "crude" model of binaural unmasking for HI listeners was not properly tested with the stimuli considered below, because those did not contain realistic ITDs. Further investigation is needed concerning this component of the model.

The internal noise considered at each ear in the model is spectrally-shaped using the tonal audiogram. The audiometric pure tone thresholds (given in dB HL) are converted into ear drum levels (in dB SPL), and the resulting levels are then interpolated to get their values at the center frequencies used in the model. The level conversion was realized by adding reference equivalent sound pressure levels for the applied THD 39 headphones [5] and nominal values for the transformation from 6 cc coupler to ear drum levels [6] to the pure tone thresholds. Within the audiometric frequency range, the thresholds in dB SPL were interpolated on a logarithmic frequency scale. For frequencies below 250 Hz and above 6 kHz, the threshold was set to the value in dB SPL[2] at 250 Hz and 6 kHz, respectively. Individual pure tone thresholds were considered separately for the left and right ears[3]. The internal noise levels are then obtained by adding a value in dB to the interpolated thresholds. This value *margin* sets the broadband level of the internal noise in dB SPL. It is a free parameter of the model, assumed to be constant across frequency and within subject group (i.e., NH or HI), but different between subject groups.

The model predictions presented here were computed using the stimuli from two related experiments [7, 8] briefly summarized in Section 3. For each condition, two minutes of the masker signal was considered and the target was represented by averaging 120 and 128 target sentences for experiments 1 and 2, respectively; after all sentences had been truncated to the duration of the shortest sentence. The masker and averaged target signals were all convolved with the impulse response of the headphones used for data collection and measured on a 4128C Bruel&Kjaer head and torso simulator. All signals were calibrated to the sound levels (dB SPL) used in the experiments.

3. Data

The experimental data and stimuli used to verify the proposed binaural model were taken from two experiments described in detail in [7] and [8]. Experiment 1 evaluated the effect of temporal masker fluctuations on SRT and SRM in NH and HI listeners [7]. Experiment 2 focused explicitly on the effect of sensation level (and thus audibility) on SRT and SRM in fluctuating noise [8].

In both experiments, SRTs were measured adaptively using BKB-like target sentences [9] in the presence of two noise-vocoded speech interferers. The noise vocoder was applied to minimize informational masking effects. It was realized with five frequency channels with a bandwidth of four critical bands each, and was applied separately to each of the two speech maskers. The target speech was unprocessed (i.e., not vocoded) and always presented from 0° azimuth, whereas the two interferers were either co-located with the target or (artificially) spatially separated. The target speech and the co-located interferers were spatialized by applying the same across-ear averaged head-related transfer function for frontal sound incidence from [10] to both ears. The spatially separated interferers were realized artificially such that one was presented to the left

[1] This value was chosen after testing this updated implementation of the model of Collin and Lavandier on different data sets from the literature. This work by Vicente and Lavandier is not published yet.

[2] It might be more appropriate to do this extrapolation on the threshold levels (dB HL) rather than on the ear drum levels (dB SPL); however, given the importance function of the speech intelligibility index (used in the model) at these very low and high frequencies, it is unlikely that model predictions would be much affected.

[3] For the modeling of experiment 1, the pure tone thresholds of the NH listeners were set to 0 dB HL.

ear and the other one to the right ear, realizing "infinite" broadband ILDs but no ITD. All stimuli were presented via equalized Sennheiser HD215 headphones and were filtered such that they had the same long-term spectrum as the target speech.

In experiment 1, the combined interferer level was set to 60 dB SPL and the level of the target speech was adjusted adaptively such that, on average, 50% of the words were correctly understood. The resulting SNR provided an estimate of the SRT. To partly compensate for the loss in audibility, the HI listeners received linear amplification according to the National Acoustic Laboratory Revised-Profound prescription formula (NAL-RP, [11]). Moreover, two uncorrelated speech-shaped noise interferers were tested in addition to the noise-vocoded speech interferers. Ten NH listeners (hearing thresholds below 15 dB HL) with a mean age of 33.1 years and ten HI listeners with a mean age of 66.9 years participated in experiment 1. All HI listeners had symmetric, mild to moderate, sloping, sensorineural hearing loss with a four frequency (0.5, 1, 2, 4 kHz) average hearing loss (4-FAHL) of 37.8 +/- 7.1 dB HL.

In experiment 2, all stimuli were audibility equalized across frequency by providing amplification (or attenuation) equivalent to the individually measured detection thresholds for speech-shaped noise filtered into nine different frequency regions. SRTs were measured for the noise-vocoded speech interferers presented at four different sensation levels (0, 10, 20 and 30 dB) relative to the individual SRTs in quiet. It should be noted that 0 dB SL corresponds to very low levels in dB SPL, in particular for the NH listeners. The level of the target is varied adaptively relative to the combined interferer level during each SRT measurement. By varying the overall level of the stimuli in this experiment, their audibility was varied. Ten NH listeners with a mean age of 23.2 years and ten HI listeners with a mean age of 70.3 years participated in experiment 2, but not all HI listeners could be tested at the higher sensation levels due to loudness tolerance issues. All HI listeners had symmetric, mild to moderate, sloping, sensorineural hearing loss with a 4-FAHL of 29.1 +/- 8.0 dB HL.

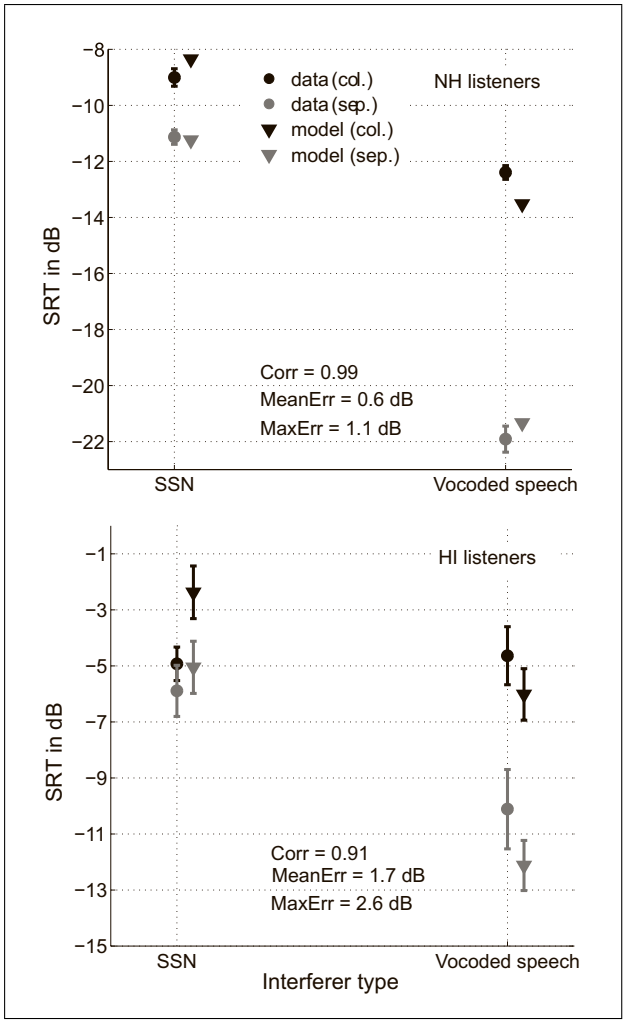

Figure 1. Mean SRTs with standard errors across NH (top panel) and HI (bottom panel) listeners measured and predicted in experiment 1. The two maskers were either speech-shaped noise (SSN) or vocoded speech, and either spatially separated (sep.) or co-located (col.) with the frontal target. A small horizontal offset has been added to the model predictions to reduce symbol overlap.

4. Predictions

Predicted differences of (inverted) binaural ratio between conditions can be directly compared to corresponding SRT differences. To compare absolute thresholds rather than relative differences, a reference needs to be chosen. For each listener considered here, the reference was the individual average SRT across conditions in the experiment. To obtain the predicted SRTs of each listener, inverted ratios were centered to this average SRT (by subtracting their mean and adding the average SRT). In other words, the individual average predicted SRT was aligned to the individual average measured SRT[4]; so that we only aimed at predicting the differences across conditions within each group of listeners and experiment (i.e., within each panel of Figures 1 and 2).

Prediction performances were evaluated in terms of Bravais-Pearson correlation between measured and predicted SRTs (*Corr*), mean absolute prediction error (*MeanErr*, absolute differences between measured and predicted SRTs averaged across conditions), and maximum absolute prediction error (*MaxErr*). The value of the free parameter *margin* of the model was chosen to minimize *MeanErr* in experiment 2, independently for each group of listeners, resulting in a *margin* of −11 dB and −22 dB for the NH and HI listeners, respectively. The same *margin* values were used for the modeling of experiment 1 (considered here for validation).

[4] In experiment 1, the stimuli, audiogram and resulting model predictions were identical for all NH listeners. The average predicted SRT across conditions was then directly aligned to the average SRT measured across listeners and conditions.

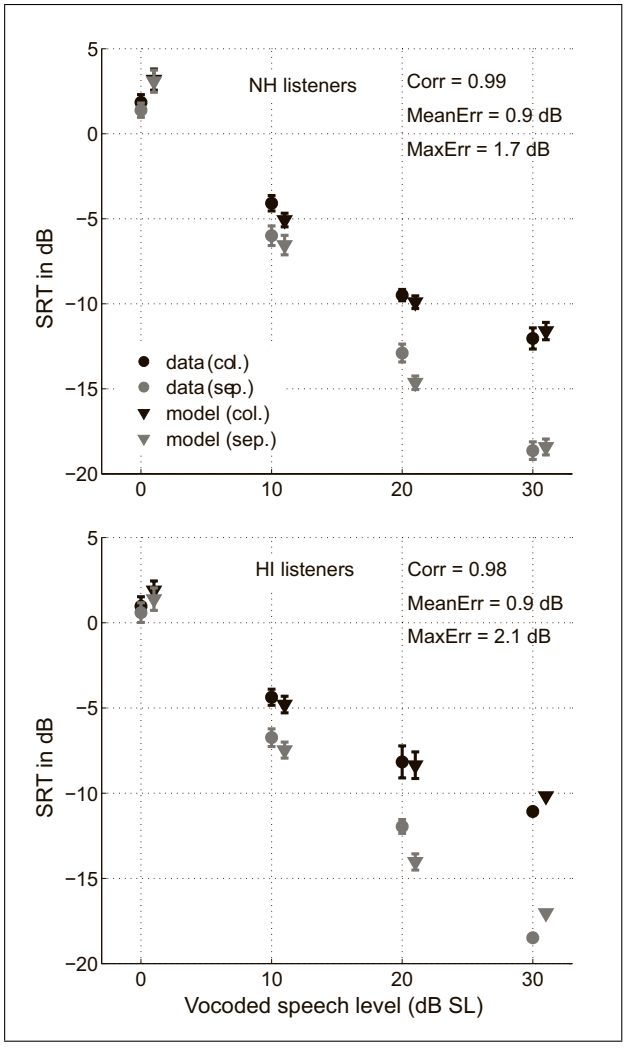

Figure 2. Mean SRTs with standard errors across NH (top panel) and HI (bottom panel) listeners measured and predicted in experiment 2 for four overall masker levels (the audibility of the target and maskers increased with overall level). The two vocoded speech maskers were either spatially separated (sep.) or co-located (col.) with the frontal target. A small horizontal offset has been added to the model predictions to reduce symbol overlap.

The measured and predicted SRTs of experiment 1 are presented in Figure 1. The model predicted closely both the SRM and the masking release associated with envelope modulations in the masker (SSN vs. vocoded speech) for the NH listeners (*MeanErr* below 1 dB; *Corr* was computed on only four points here so it should be considered with caution). Predictions were less accurate for the HI listeners (even if *MeanErr* remained below 2 dB). In particular, the model overestimated the SRM for the SSNs. This SRM was not present in the data. As a result, the model also overestimated the effect of the masker envelope modulations in the co-located condition: the model predicts an advantage that is also not apparent in the data for the HI listeners.

Figure 2 presents the mean SRTs measured in experiment 2 for the NH (top panel) and HI (bottom panel) listeners, along with the model predictions for each group. SRTs are plotted as a function of overall masker level (increasing level corresponds to increasing audibility for both target and maskers). The model described accurately both the decrease of SRT and the increase of SRM with increasing audibility for both groups of listeners (*Corr* above 0.98 and *MeanErr* below 1 dB).

5. Discussion

While tested on data measured with diotic and dichotic stimuli reproduced over headphones, the binaural model proposed here was able to predict rather accurately the release from masking due to better-ear glimpsing in the presence of two maskers, the dip listening advantage associated with envelope modulations in these maskers, and the effect of audibility on both SRTs and better-ear glimpsing, for NH and HI listeners. Predictions were less accurate for the HI listeners in experiment 1 that was not used to define the value of the free parameter of the model. Prediction performance was at least as good as for previous binaural NH models [2, 4], with a *MeanErr* between 0.6 and 1.7 dB. When stimuli levels are well above hearing thresholds, the proposed model is equivalent to the one of Collin and Lavandier [2]. It was the case for the NH listeners in experiment 1 and the similar prediction performance obtained highlights the backward compatibility of the model. Even if the better-ear glimpsing component of the model seems validated by the first predictions presented here, the model needs to be further tested using stimuli with realistic ILDs and ITDs. In particular, the binaural unmasking component of the model relying on ITDs could not be tested here.

The free parameter *margin* – used to obtain the model internal noise levels from the tonal audiograms – had to be set to different values for the NH and HI listeners. This is an important limitation of the model. When considering a panel of listeners with increasing hearing losses, it would be more relevant to be able to use a single model for all listeners, so that in the future *margin* would at least need to be made dependent on the degree of hearing loss. The fact that it is not the case in the current model might explain why less accurate predictions were obtained for the HI listeners of experiment 1 that had a larger average hearing loss (4-FAHL) than the HI listeners of experiment 2, which were the listeners considered when defining *margin*.

The difference in *margin* obtained for the NH and HI listeners could reflect potential effects of reduced spectral and temporal resolutions for the HI listeners, but also additional effects of cognitive differences between the two groups due to the age confound (i.e., young NH vs. old HI). In a different modeling framework (at least in terms of implementation, even if the concept of the present model is quite similar), Beutelmann *et al.* model the effects of audibility by adding independent internal noises at each ear. The internal noise is also spectrally-shaped using the tonal audiogram and its levels are set 1 dB [1] or 4 dB [4] above the audiometric thresholds. These values are positive and

much smaller in magnitude than the offsets of −11 dB and −22 dB used here. Even if the differences in model implementation might explain part of this discrepancy, more investigations are needed concerning this important parameter of the proposed model. It should be noted that, apart from the use of a different *margin*, the current model is identical for NH and HI listeners (e.g., in terms of spectral and temporal resolutions). This might need further refinement as well.

Importantly, prediction performances were quite similar for both groups of listeners. While using identical parameters for NH and HI predictions, the binaural model proposed by Beutelmann *et al.* could predict well SRTs measured in the presence of one SSN in different rooms [1]. The SRM was generally overestimated for HI listeners, but the difference in *MeanErr* between NH and HI listeners was only 0.5 dB. In the presence of an envelope-modulated noise [4], the predictions were less accurate for the HI compared to the NH listeners (*Corr* in the range 0.59–0.80 and 0.80–0.93, *MeanErr* of 4 and 3 dB, respectively).

Only averaged predictions across listeners were presented here. The model can be applied to predict SRTs for individual listeners, but care should be taken to assure that these individual SRTs are not influenced by the potential confounding effect of the sentence material, which needs to be counterbalanced across conditions (as it was the case for the averaged SRTs considered here). The proposed model could be a useful tool to investigate individual differences between HI listeners in the future.

Acknowledgement

The international mobility of ML at Macquarie University/NAL was funded by ENTPE, Macquarie University and CeLyA (ANR-10-LABX-0060/ANR-11-IDEX-0007). The participation of ML to ISH 2018 was funded by la Fondation pour l'Audition.

References

[1] R. Beutelmann, T. Brand: Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am. **120** (2006) 331–342.

[2] B. Collin, M. Lavandier: Binaural speech intelligibility in rooms with variations in spatial location of sources and modulation depth of noise interferers. J. Acoust. Soc. Am. **134** (2013) 1146–1159.

[3] H. Glyde, S. Cameron, H. Dillon, L. Hickson, M. Seeto: The effects of hearing impairment and aging on spatial processing. Ear Hear. **34** (2013) 15–28.

[4] R. Beutelmann, T. Brand, B. Kollmeier: Revision, extension, and evaluation of a binaural speech intelligibility model. J. Acoust. Soc. Am. **127** (2010) 2479–2497.

[5] 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. International Organization for Standardization, Geneva (1994).

[6] R. A. Bentler, C. V. Pavlovic: Transfer functions and correction factors used in hearing aid evaluation and research. Ear Hear. **10** (1989) 58–63.

[7] B. Rana, J. M. Buchholz: Better-ear glimpsing at low frequencies in normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am. **140** (2016) 1192–1205.

[8] B. Rana, J. M. Buchholz: Effect of audibility on better-ear glimpsing as a function of frequency in normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am. **143** (2018) 2195–2206.

[9] J. Bench, A. Kowal, J. Bamford: The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. Brit. J. Audiol. **13** (1979) 108–112.

[10] S. Cameron, H. Dillon: Development of the listening in spatialized noise-sentences test (LISN-S). Ear Hear. **28** (2007) 196–211.

[11] H. Dillon: Hearing aids. 2nd ed. Boomerang Press, Sydney, 2012, 290–297.