BINAURAL PREDICTION OF SPEECH INTELLIGIBILITY FOR HEARING-IMPAIRED AND NORMAL-HEARING LISTENERS IN THE PRESENCE OF NON-STATIONARY NOISE AND REVERBERATION

Raphael Cueille[1-2*] **Mathieu Lavandier**[1] **Nicolas Grimault**[2]
[1] Univ Lyon, ENTPE, LTDS, UMR 5513, rue M. Audin, 69518 Vaulx-en-Velin, France
[2] CRNL, UMR CNRS 5292, Univ. Lyon 1, 50 av T. Garnier, 69366 Lyon Cedex 07, France

ABSTRACT

A binaural model was developed to predict speech intelligibility for hearing-impaired and normal-hearing (NH) listeners in modulated and stationary noises, at low and high levels of reverberation. It combines two existing models. First, the Leclère et al. (2015) model takes binaural room impulse responses (BRIRs) as input to compute binaural useful-to-detrimental (U/D). It accounts for the temporal smearing of the target speech in reverberant environments, but only works with stationary noises for NH listeners. Second, the Vicente et al. (2020) model takes the speech and noise signals at the ears as inputs and accounts for hearing loss by creating an internal noise depending on the listener; but it cannot account for the detrimental effect of reverberation smearing the target speech. The new model takes the audiogram, BRIRs and ear signals as inputs to account for both temporal smearing and hearing loss and works for both stationary and modulated noises. It was tested by comparing its predictions with speech reception thresholds measured in seven experiments. The predictions obtained were as accurate as those obtained with the two previous models that cannot be used to predict all datasets. Several methods to determine the U/D limit were compared.

Keywords: *speech intelligibility, hearing impairment, binaural hearing, reverberation, models*

[*] ***Corresponding author***: raphael.cueille@entpe.fr

1. INTRODUCTION

The spatial release from masking (SRM) is defined as the ability to benefit from spatially separated target and masker sources (Bronkhorst & Plomp, 1988). It is essential to speech intelligibility in binaural environments. It can be divided in two mechanisms, better-ear listening and binaural unmasking. Better-ear listening originates from interaural level differences between the target and masker signals at the two ears, while binaural unmasking relies on interaural time differences between the target and the masker. SRM is reduced for hearing-impaired (HI) listeners, mainly because of their elevated hearing thresholds (Glyde et al., 2011).

Reverberation has several effects on speech intelligibility. It temporally smears the target speech (Houtgast & Steeneken, 1985), reduces the efficiency of the "dip listening" mechanism, that is to say the intelligibility gain due to the masker's temporal modulations (Festen & Plomp, 1990), and decreases the SRM (Plomp, 1976). HI listeners are also affected by these effects, at least as much as NH listeners (Cueille et al., 2022; Xia et al., 2018).

Most binaural speech intelligibility models can predict SRM and the effect of reverberation on SRM (Lavandier & Best, 2020). There exists some models which can also account for the effect of hearing loss (Beutelmann & Brand, 2006; Vicente et al., 2020), but they usually cannot predict the influence of the temporal smearing of the target speech caused by reverberation. Other models can predict the effect of the temporal smearing of the target, but only with NH listeners (Leclère et al., 2015). To the best of our knowledge, the only model which can account for both the effect of hearing loss and the effect of the temporal smearing of the target speech has never been tested with HI listeners (Rennies et al., 2011).

The main objective of this study was to develop a binaural model able to predict speech intelligibility for hearing-impaired listeners in binaural and reverberant conditions. In order to do so, the models of Leclère et al., 2015 and Vicente et al., 2020 (called here leclere2015 and vicente2020) were combined.

2. MODEL DESCRIPTION

2.1 Original models

Both the leclere2015 and vicente2020 models are based on the same model developed by Lavandier & Culling, 2010 (called lavandier2010) here. This model takes the BRIRs recorded at the ears of the listener as inputs. It predicts the better-ear listening and the binaural unmasking by frequency bands, before being averaged with a SII weighting. The sum of the resulting better-ear listening and binaural unmasking is defined as a binaural ratio. Differences between predicted binaural ratios can be compared to differences between SRTs. This model can predict SRM and the effect of reverberation on SRM. However it cannot predict the effect of the temporal smearing of the target speech and of the dip listening. Furthermore, it only works with NH listeners.

The leclere2015 model is an extension of the lavandier2010 model, with a similar structure. Its main difference is that the target BRIR is separated in an early and a late part. The early part is considered as useful to speech intelligibility, while the late part is considered as detrimental and is concatenated with the masker BRIR. This allows to consider the effect of the temporal smearing of the target speech caused by reverberation. However it only works with NH listeners and cannot predict the dip listening benefit.

The vicente2020 model takes the target and masker signals at the ears of the listener as inputs. The better-ear listening and binaural unmasking are computed by time bands in order to take into account the effect of the masker's temporal modulations. Another major difference is that the vicente2020 model also takes the listener's audiogram as input. An internal noise is created based on the audiogram and the external noise level (that is to say, the masker level). The masker sound level at each ear is taken as the highest level between the internal noise level and the external noise level. This allows to predict the influence of hearing loss. However, this model cannot predict the effect of the temporal smearing of the target speech.

2.2 Proposed model

The proposed model combines the leclere2015 and vicente2020 models. It consists in applying the vicente2020 model on different inputs. The original speech and noise materials, the BRIRs recorded at the ears, and the listener's audiograms are taken as inputs. The stimuli used in the experiment can then be recreated by convoluting the speech and noise materials with the BRIRs. However, the target BRIR is separated in an early and a late part before convolution. The speech material is convolved with both the early and late target BRIR. The early target signal is considered as the useful signal and is taken as input in the vicente2020 model. The late target signal is summed with the masker signal (that is to say, the noise material convolved with the masker BRIR), and this sum is taken as input in the vicente2020 model as the detrimental signal.

2.3 BRIR separation parameters

The target BRIR separation part is done with linear windows. Both the early and late target BRIRs consist of a flat part as well as a linear decreasing (for the early BRIR) or increasing (for the late BRIR) part. Two parameters must be determined. The early-late-limit (ELL) is defined as the duration of the flat part of the early BRIR. The decay duration (DD) is defined as the duration of the linear decrease of the early BRIR. Leclère et al. (2015) showed that the leclere2015 model gave more accurate predictions when these parameters were room dependant.

Four early-late separation methods were used with the proposed model. A room independent separation proposed by Leclère et al. (2015) was tested, with an ELL and a DD fixed at 30 ms and 25 ms respectively. A separation based on the work of Lindau et al., 2012 was also tested, where the ELL was considered as the perceptual mixing time while the DD remained fixed at 25 ms. The third version was based on a linear regression between the ELL and the

IACC, developed by Kokabi et al., 2018. The fourth version was an attempt to link the ELL with a short-term IACC, based on the work of (Rennies et al., 2019).

3. DATASETS

Six datasets were used to test the models predictions. (Collin & Lavandier, 2013 - Exp.1-3-4; Cueille et al., 2022; Lavandier et al., 2012 - Exp.1; Lavandier & Culling, 2008 - Exp.3). The Lavandier et al., 2012 dataset allowed to see the SRM and the effect of reverberation on SRM. The Lavandier & Culling, 2008 dataset allowed to see the influence of reverberation on SRM and the temporal smearing of the speech. The Collin & Lavandier, 2013 datasets allowed to see the effect of reverberation on dip listening and SRM. The Cueille et al., 2022 allowed to predict the effect of reverberation on dip listening as well as the temporal smearing of the target in monaural conditions. It also was the only dataset with HI listeners.

4. RESULTS

The four model versions gave similar performances with all datasets. Pearson’s correlation coefficients between measured and predicted SRTs ranged between 0.8 and 0.99. Mean absolute errors ranged between 0.2 dB and 1.5 dB. Largest absolute errors ranged between 1.5 and 3.7 dB. The versions based on the work of Kokabi et al., 2018 and on the short-term IACC gave slightly less accurate predicions with the datasets with high target temporal smearing, that is to say the experiment 3 of Collin & Lavandier, 2013 and Lavandier & Culling, 2008. The dataset from Cueille et al., 2022 was slightly less accurately predicted than the others.

5. DISCUSSION

The four versions of the proposed model give accurate predictions with all datasets and are at least able to predict the trends of the studied effects. The slightly less accurate predictions of the Kokabi version might be explained by the fact that the ELL-IACC linear regression was developed for low and intermediary reverberation levels and was not tested at high reverberation levels. It appears that using a room independent ELL separation might be more convenient, as its implementation is simpler, and the predictions are as accurate as the other versions.

The proposed model gives better predictions than the vicente2020 and the leclere2015 models. The predictions of the vicente2020 model are accurate only with datasets with no temporal smearing, while the leclere2015 model is inaccurate with datasets with a temporally modulated masker and/or HI listeners.

Overall, the proposed model appears to work correctly. However, the datasets used here only allow to test the different mechanisms (SRM, target temporal smearing, dip listening…) separately. Furthermore, the only dataset with HI listeners has monaural conditions with no SRM. It could be interesting to test this model with HI listeners in binaural conditions.

6. REFERENCES

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

Bronkhorst, A. W., & Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 9.

Collin, B., & Lavandier, M. (2013). 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.

Cueille, R., Lavandier, M., & Grimault, N. (2022). Effects of reverberation on speech intelligibility in noise for hearing-impaired listeners. *Royal Society Open Science*, *9*(8), 210342.

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

Glyde, H., Hickson, L., Cameron, S., & Dillon, H. (2011). Problems Hearing in Noise in Older Adults : A Review of Spatial Processing Disorder. *Trends in Amplification*, *15*(3), 116-126.

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

Kokabi, O., Brinkmann, F., & Weinzierl, S. (2018). Segmentation of binaural room impulse responses for speech intelligibility prediction. *The Journal of the Acoustical Society of America*, *144*(5), 2793-2800.

Lavandier, M., & Best, V. (2020). Modeling Binaural Speech Understanding in Complex Situations. In J. Blauert & J. Braasch (Éds.), *The Technology of Binaural Understanding* (p. 547-578). Springer International Publishing.

Lavandier, M., & Culling, J. F. (2008). 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.

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

Lavandier, M., Jelfs, S., Culling, J. F., Watkins, A. J., Raimond, A. P., & Makin, S. J. (2012). Binaural prediction of speech intelligibility in reverberant rooms with multiple noise sources. *The Journal of the Acoustical Society of America*, *131*(1), 218-231.

Leclère, T., Lavandier, M., & Culling, J. F. (2015). 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.

Lindau, A., Kosanke, L., & Weinzierl, S. (2012). Perceptual Evaluation of Model- and Signal-Based Predictors of the Mixing Time in Binaural Room Impulse Responses. *Journal of the Audio Engineering Society*, *60*(11), 887-898.

Plomp, R. (1976). Binaural and Monaural Speech Intelligibility of Connected Discourse in Reverberation as a Function of Azimuth of a Single Competing Sound Source (Speech or Noise). *Acta Acustica united with Acustica*, *34*(4), 200-211.

Rennies, J., Brand, T., & Kollmeier, B. (2011). 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.

Rennies, J., Warzybok, A., Brand, T., & Kollmeier, B. (2019). Measurement and Prediction of Binaural-Temporal Integration of Speech Reflections. *Trends in Hearing*, *23*.

Vicente, T., Lavandier, M., & Buchholz, J. M. (2020). 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.

Xia, J., Xu, B., Pentony, S., Xu, J., & Swaminathan, J. (2018). Effects of reverberation and noise on speech intelligibility in normal-hearing and aided hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *143*(3), 1523-1533.