DOI: 10.61782/fa.2023.0102

TOWARDS A BINAURAL HEARING AID SPEECH PERCEPTION INDEX (HASPI): PREDICTIONS OF ANECHOIC SPATIAL RELEASE FROM MASKING FOR NORMAL-HEARING LISTENERS

Mathieu Lavandier[1*] **James M. Kates**[2] **Kathryn H. Arehart**[2]

[1] Univ Lyon, ENTPE, Ecole Centrale de Lyon, CNRS, LTDS, UMR5513, 69518 Vaulx-en-Velin, France

[2] Department of Speech Language and Hearing Sciences, University of Colorado, Boulder, CO 80309, USA

ABSTRACT

The Hearing Aid Speech Perception Index (HASPI v2) is a speech intelligibility metric allowing predictions for speech degraded by additive noise, reverberation, spectral changes, and nonlinear distortion. It has been developed using monaural sentences presented over headphones. HASPI can also predict the effects of hearing impairment on intelligibility by incorporating a model of the auditory periphery that can represent impaired and normal hearing. HASPI is an intrusive metric that compares a degraded signal to a clean reference. Its current version has been developed to predict percent correct for English sentences measured at positive signal-to-noise ratios (SNRs). The aims of the present study were to test (1) a model back-end so that HASPI could be used to predict percent correct for Danish words measured at negative SNRs (2) a binaural front-end to predict the spatial release from masking that can be provided by binaural hearing. At this stage, model predictions were compared to intelligibility scores from normal-hearing listeners, measured in anechoic conditions for a frontal (Danish) speech target, in the presence of a single stationary speech-shape noise tested at ten azimuths around the listener and six SNRs varying across conditions between -1.5 and -26.5 dB [1-2].

*Corresponding author: mathieu.lavandier@entpe.fr.

Keywords: *speech intelligibility, binaural hearing, models.*

1. INTRODUCTION

The Hearing Aid Speech Perception Index (HASPI v2; [3]) is a monaural metric allowing to predict speech intelligibility when speech is affected by noise, reverberation, and nonlinear distortion (e.g. due to hearing-aid signal processing). Because HASPI incorporates a model of the normal and impaired auditory peripheries, it can also account for effects associated with hearing loss. However, because it is monaural, it cannot account for the advantage provided by binaural hearing, and the spatial release from masking it produces when the target speech and interfering noise sources are at different positions [4].

The Modified Binaural Short-Time Objective Intelligibility (MBSTOI; [2]) is a binaural metric able to predict spatial release from masking, along with the effects of reverberation and nonlinear distortion, but it cannot account for the effect of hearing loss, so that predictions are limited to normal-hearing listeners.

The aim of the present study was to combine the two models to benefit from their respective advantages.

2. PERCEPTUAL DATA USED TO TEST THE PROPOSED MODEL

As a first step to develop a binaural HASPI, we wanted this metric to be able to predict spatial release from masking, independently from the effect of reverberation, nonlinear distortion and hearing loss. Preliminary model versions were tested on a very simple dataset highlighting strong spatial release from masking: the dataset D1 measured by Andersen et al. (2016, 2018) and used to develop MBSTOI: intelligibility scores from normal-hearing listeners, measured in anechoic conditions for a frontal (Danish) speech target, in the presence of a single stationary speech-shape noise tested at ten azimuths around the listener and six SNRs varying across conditions between -1.5 and -26.5 dB [1-2]. One of the tested azimuths was 0 degree, corresponding to a frontal noise masker co-located with the target speech. This condition was used as a reference in which there is no spatial release from masking and no advantage of binaural hearing.

Model predictions were compared to the intelligibility scores averaged across listeners.

3. MODELLING APPROCH

Both HASPI and MBSTOI are correlation models [5] that compares the degraded noisy speech signal to a clean speech reference, so that they use the same model inputs.

HASPI has been developed to predict percent correct for English sentences measured at positive signal-to-noise ratios (SNRs). The first step was to develop a model back-end so that HASPI could be used to predict percent correct for Danish words measured at negative SNRs.

HASPI computes ten features when comparing the degraded noisy speech signal to the clean speech reference, these features being then mapped to percent correct. A different mapping function was developed so that it could be used for predictions at negative SNRs with Danish speech material. The idea was to develop this mapping by considering only conditions in which binaural hearing was not involved, so that the binaural front-end of the model could then be tested independently from this mapping to intelligibility scores. This was done by considering only the co-located condition in the dataset, along with other co-located conditions using the same speech material, provided by datasets D2 and D5 from Andersen et al. (2016, 2018), which do not involve binaural hearing but comprise other tested SNRs, and also modulated noise and nonlinear processing [2].

Our second step was to develop a binaural front-end to be used with HASPI. For that, we are currently modifying MBSTOI so that it can produced binaural-enhanced signals such as done in previous binaural speech intelligibility model (e.g. [6]).

4. RESULTS

The first step concerning the back-end of the model was achieved: the ten HASPI features still highlight variations at the low SNRs considered in the Danish datasets, so that they could be successfully mapped to the intelligibility scores in the co-located conditions.

Predictions are still currently being computed and compared to the data for different versions of the binaural front-end. They will be presented at the conference.

5. REFERENCES

[1] A. H. Andersen, Z.-H. Tan, J. M. de Haan, and J. Jensen: "Predicting the intelligibility of noisy and nonlinearly processed binaural speech," *IEEE Trans. Audio, Speech, Language Process.* 24 (11), 1908-1920, 2016.

[2] A. H. Andersen, Z.-H. Tan, J. M. de Haan, and J. Jensen: "Refinement and validation of the binaural short time objective intelligibility measure for spatially diverse conditions," *Speech Communication* 102, 1-13, 2018.

[3] J. M. Kates and K. H. Arehart: "The hearing-aid speech perception index, version 2", Speech Comm. 131, 35-46, 2021.

[4] J.F. Culling, M. Lavandier: "Binaural unmasking and spatial release from masking". In: R.Y. Litovsky, M.J. Goupell, A.N. Popper, R.R. Fay, Eds. Binaural Hearing, Vol. 73 of Springer Handbook of Auditory Research, Switzerland: Springer Nature, 209–241, 2021.

[5] M. Lavandier and V. Best: "Modeling binaural speech understanding in complex situations," in The Technology of Binaural Understanding, edited by J. Blauert and J. Braasch (Springer Nature, Cham, Switzerland), pp. 547–578, 2020.

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