

# Binaural Speech Intelligibility in Noise and Reverberation: Prediction of Group Performance for Normal-hearing and Hearing-impaired Listeners

Trends in Hearing

Volume 29: 1–21

© The Author(s) 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/23312165251344947

journals.sagepub.com/home/tia



Raphael Cueille<sup>1,2</sup> and Mathieu Lavandier<sup>1</sup>

## Abstract

A binaural model is proposed to predict speech intelligibility in rooms for normal-hearing (NH) and hearing-impaired listener groups, combining the advantages of two existing models. The *ledere2015* model takes binaural room impulse responses (BRIRs) as inputs and accounts for the temporal smearing of the speech by reverberation, but only works with stationary noises for NH listeners. The *vicente2020* model takes the speech and noise signals at the ears as well as the listener audiogram as inputs and accounts for modulations in the noise and hearing loss, but cannot predict the temporal smearing of the speech by reverberation. The new model takes the audiogram, BRIRs and ear signals as inputs to account for the temporal smearing of the speech, the masker modulations and hearing loss. It gave accurate predictions for speech reception thresholds measured in seven experiments. The proposed model can do predictions that neither of the two original models can make when the target speech is influenced by reverberation and the noise has modulations and/or the listeners have hearing loss. In terms of model parameters, four methods were compared to separate the early and late reverberation, and two methods were compared to account for hearing loss.

## Keywords

speech intelligibility, auditory modelling, binaural hearing, reverberation, hearing loss

Received: May 14, 2024; revised: April 9, 2025; accepted: May 1, 2025

## Introduction

Hearing-impaired (HI) listeners often complain about difficulties to understand speech in noisy and reverberant rooms. The aim of the present study was to develop a model able to predict the effect of audibility loss on speech intelligibility in such environments.

One key mechanism to understand speech in noisy environments is spatial release from masking (SRM), the ability to benefit from a difference in position between the target speech and noise masker sources (Bronkhorst & Plomp, 1988). It can be separated in two mechanisms, better-ear listening and binaural unmasking. Better-ear listening originates from differences of interaural level differences (ILDs) between the target and masker signals at the two ears, which results in differences of signal-to-noise ratio (SNR) at the two ears, allowing us to benefit from the better SNR between the two ears to better understand the target speech. Binaural unmasking is dependent on the masker interaural coherence and on the differences of interaural time

differences (ITD) produced by the target and masker. It can be modeled by assuming that the listener is able to equalize the masker at the two ears, and partially cancel it to improve intelligibility (Durlach, 1963). SRM is usually reduced for HI listeners (Glyde et al., 2011). First, their elevated hearing thresholds decrease the audibility of the speech signal, which can limit the benefit of better-ear listening (since the speech signal at the better ear may not be audible). Second, they have a reduced sensitivity to the temporal fine structure of sounds (Füllgrabe & Moore, 2017), which

<sup>1</sup>ENTPE, Ecole Centrale de Lyon, CNRS, LTDS, Vaulx-en-Velin, France

<sup>2</sup>CRNL, UMR CNRS 5292, Univ. Lyon 1, Lyon, France

### Corresponding author:

Raphael Cueille, ENTPE, Ecole Centrale de Lyon, CNRS, LTDS, UMR5513, 69518 Vaulx-en-Velin, France.

Email: raphael.cueille@cerema.fr

Data Availability Statement included at the end of the article



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

might limit their ability for binaural unmasking (Vicente et al., 2021).

Room reverberation has several detrimental effects on speech intelligibility in noise. It temporally smears the target speech, which degrades its intelligibility (Houtgast & Steeneken, 1985). Reverberation also temporally smears the masker. This can be detrimental to intelligibility when the masker is modulated in amplitude, as it reduces the intelligibility benefit coming from listening in the “dips” of the masker, also called dip listening benefit (Festen & Plomp, 1990; Miller & Licklider, 1950). SRM is also reduced by reverberation (Plomp, 1976). Sound reflections travelling around the head reduce ILDs, impairing better-ear listening. These reflections also decorrelate the masker at the two ears, reducing binaural unmasking (Lavandier & Culling, 2008). The sensitivity of HI listeners to reverberation has been a matter of debate. They seem to be as much affected as normal-hearing (NH) listeners by the effects of reverberation in monaural conditions (Cueille et al., 2022). There is some evidence that this is also the case in binaural conditions (Helfer & Wilber, 1990; Xia et al., 2018).

Several binaural speech intelligibility models have been proposed in the literature (for a review, see Lavandier & Best, 2020). Most predict SRM and the effects of reverberation on SRM. Some of these models can also predict the degradation of speech intelligibility caused by reverberation temporally smearing the target speech but only for NH listeners (Chabot-Leclerc et al., 2016; Leclère et al., 2015; van Wijngaarden & Drullman, 2008). Other models allow predictions for HI listeners (Beutelmänn & Brand, 2006; Beutelmänn et al., 2010; Vicente et al., 2020), but cannot account for the temporal smearing of the target speech by reverberation. Rennie et al. (2011) developed three extensions of the model proposed by Beutelmänn et al. (2010). Those extensions allow to predict speech intelligibility for HI listeners and temporally smeared speech. Rennie et al. (2019) found that one can create particular conditions where the late reflections of the target speech can be useful to intelligibility for NH listeners, such as when these reflections have more energy than the direct sound, or when they carry a different interaural phase difference to the one of the masker. HI listeners seem to be less able to “ignore” the direct sound than NH listeners in these conditions (Rennie et al., 2022). Rennie et al. (2019) proposed a fourth model extension to account for these mechanisms. However, to the best of our knowledge, the model versions from Rennie et al. (2011) and Rennie et al. (2019) were only tested for NH listeners and a stationary noise masker.

Andersen et al. (2016, 2018) proposed binaural models based on the models of Taal et al. (2011) and Beutelmänn et al. (2010). The models of Andersen et al. (2016) and Andersen et al. (2018) can predict the performances of NH listeners in reverberant conditions, and the effect of nonlinear speech processing similar to those used in hearing aids. However, nothing is currently implemented so that they could

predict the effect of hearing loss. During the Clarity Challenge (Barker et al., 2022), several binaural models using a machine learning approach were tested on datasets involving HI listeners and signals processed by different speech enhancement algorithms. The model from Schädler et al. (2018) follows a similar approach. In principle, these models are able to do predictions for HI listeners wearing hearing aids in rooms. However, some of the more accurate models of the Clarity Challenge as well as the model from Schädler et al. (2018) require being trained with speech and conditions very similar to the evaluation material beforehand.

A new model was developed here, combining the models of Vicente et al. (2020) and Leclère et al. (2015). It is able to predict the mean speech reception thresholds (SRTs, the SNRs for 50% intelligibility) of NH and HI listener groups, for datasets involving nonstationary noises in binaural and reverberant conditions, something that the two original models cannot do. Using seven datasets described in more details below, this model is shown to account for SRM and the influence of reverberation on SRM, dip listening and the influence of reverberation on dip listening, temporal smearing of the target speech by reverberation, and hearing loss. These perceptual mechanisms and reverberation effects of reverberation were tested separately and in combinations to better highlight the ability of the model to predict each of the multiple effects at play (in isolation and when combined).

## Description of the Models

### Original Models

Both the Vicente et al. (2020) and Leclère et al. (2015) models (*vicente2020* and *leclere2015* in the following and in Lavandier et al. (2022)) are based on a model developed by Lavandier and Culling (2010). This model allows to predict better-ear listening and binaural unmasking advantages by frequency bands, taking the signals recorded at the ears of the listener as inputs. Instead of the target speech, the target signal consists of a speech-shaped noise with the long-term characteristics of the target speech (magnitude spectrum and ITD), obtained by averaging a large number of target sentences. This is done in order to avoid that the short pauses between words lead to very low SNRs. Better-ear listening is implemented by evaluating the SNR at each ear by frequency band. The maximum SNR across ears is integrated across frequency using a Speech Intelligibility Index (SII) weighting (ANSI S3.5, 1997) to obtain a broadband target-to-interferer ratio. The binaural unmasking advantage is computed in each frequency band following a formula proposed by Culling et al. (2005), and integrated across bands using the SII weightings. The binaural unmasking advantage and better-ear ratio are summed in order to obtain a binaural ratio. The binaural ratio is an SNR value in dB which reflects speech intelligibility. A higher binaural ratio represents a better speech intelligibility. Differences in binaural ratios can be compared

to differences in SRTs. To be compared to absolute SRTs, an SNR value needs to be chosen, which is often the average SRT measured in the experiment (Lavandier et al., 2022).

The *leclere2015* model is an extension of the Lavandier and Culling (2010) model based on the concept of useful-to-detrimental ratio (Lochner & Burger, 1964). It takes the binaural room impulse responses (BRIRs) at the ears of the listener as input. The target BRIR is separated into an early and a late part. The early part is considered as useful for speech intelligibility, while the late part is considered as detrimental and is concatenated with the masker BRIR. The early-late separation can be done using several temporal window shapes: rectangular, linear, sigmoid (Leclère et al., 2015). The *leclere2015* model uses “linear” windows: the early window starts with a constant amplitude of 1, followed by a linear decrease to 0, while the late window starts at 0 and then increases linearly to 1, so that the two windows sum to 1 at each instant. These linear windows are defined by two parameters. The early-late limit (ELL) defines the duration of the initial constant portion of the early window. The decay duration defines the duration of the linear decrease of this early window. Even if Leclère et al. (2015) found that the best model performances were obtained when these two parameters were room-dependent, *leclere2015* uses room-independent ELL and decay duration, fixed to 30 and 25 ms, respectively (Leclère et al., 2015). These values were shown to be the best compromise solution when these parameters are fixed across rooms/positions. This model accounts for the detrimental effects of reverberation on the target speech and on the SRM for stationary noises. However, because it does not take the masker signal and the listener’s audiogram as inputs, it cannot predict the effects of masker modulations and hearing impairment.

The *vicente2020* model is a different extension of the Lavandier and Culling (2010) model with two main improvements: A short-term analysis to take into account the effect of masker modulations (Collin & Lavandier, 2013), and an internal noise modelling the effect of hearing impairment (Vicente et al., 2020). Its structure is shown on the right part of Figure 1. It takes the target and masker signals at the ears of the listener as inputs. The target signal is a speech-shaped noise with the long-term characteristics of the target speech. The masker signal is segmented in time frames and the better-ear ratio and binaural unmasking advantage are calculated by frequency bands in each time frame (Collin & Lavandier, 2013). The time frames have durations of 24 and 300 ms for the better-ear ratio and binaural unmasking advantage, respectively. The use of longer time frames for the binaural unmasking advantage accounts for binaural sluggishness (Vicente & Lavandier, 2020).

An internal noise spectrum is also implemented at each ear, with spectrum levels depending on the listener audiogram and the external sound broadband level. Vicente et al. (2020) approximated this external sound level with the masker level, assuming that the broadband SNR is below 0 dB. The hearing

loss of the listeners is split into proportions to reflect the contributions of outer hair cell and inner hair cell loss. The maximum value allowed for the outer hair cell loss is 57.6 dB (Moore & Glasberg, 2004) and loss above this value contributes to the estimated inner hair cell loss. The outer hair cell and inner hair cell losses are interpolated at the model’s center frequencies between 250 and 8000 Hz, and extrapolated otherwise. The level of the internal noise  $N_{int}$  is then calculated with the following formula:

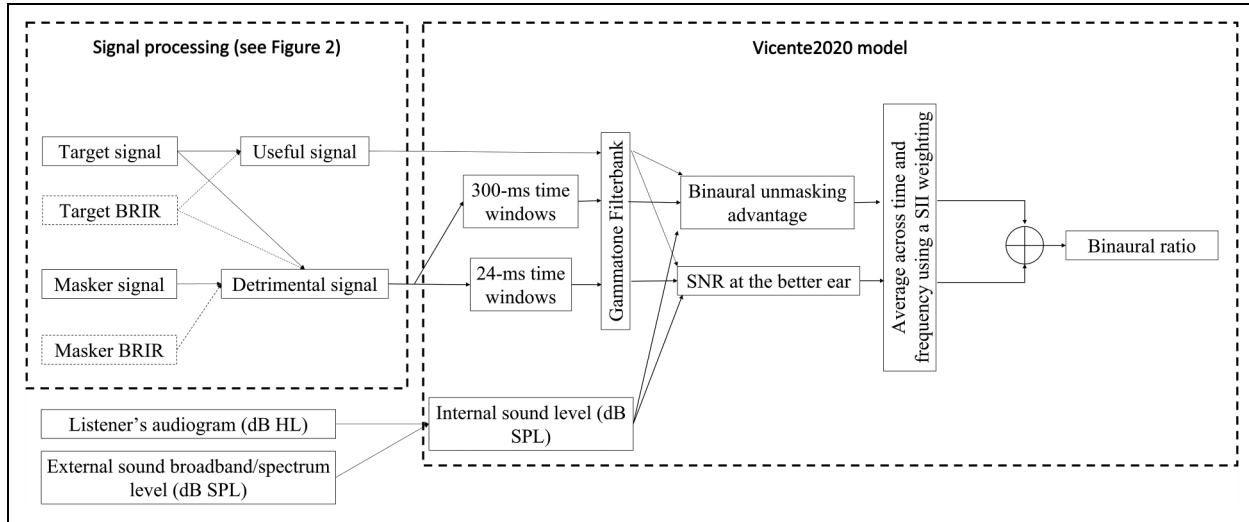
$$N_{int} = \Gamma(n) + \eta T(n) + 10 \log_{10}(10^{B/10} + 10^{[N - N_{lim} + (1 - \eta)T(n)]/10}), \quad (1)$$

where  $n$  is the center frequency of the  $n$ th frequency band and  $N$  is the long-term broadband level of the external masker average across ears.  $T(n)$  refers to the standard pure-tone audiogram in dB HL.  $\Gamma(n)$  is the transformation to convert the hearing loss from dB HL to dB SPL.  $B$ ,  $N_{lim}$  and  $\eta$  are frequency-independent parameters. They were systematically varied by Vicente et al. (2020) for experiments involving HI listeners in order to determine their value.  $B$  is the internal noise floor and is fixed at -10 dB SPL.  $N_{lim}$  is the noise level at which the internal noise level starts to increase with the external noise level. It is fixed at 83 dB SPL.  $\eta$  defines the percentage of hearing threshold elevation due to OHC loss and is fixed at 0.7.

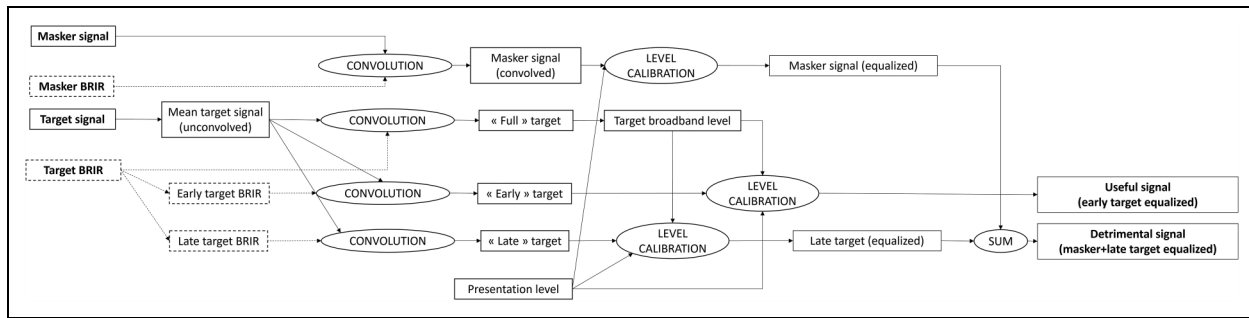
The masker and target signals need to be calibrated to the external sound level. When calculating the better-ear SNR, the SNR at each ear is computed using the higher level between the internal noise and the external masker. The binaural unmasking advantage is computed as done by Lavandier and Culling (2010), except that for each frequency band, the binaural unmasking advantage is set to zero as soon as the target or masker level is below the internal noise level at one ear. The *vicente2020* model allows to predict SRM for NH and HI listeners. However, because it does not consider the early and late parts of the target separately (and thus considers them both useful), it cannot account for the temporal smearing of the target by reverberation.

## Proposed Model

**Model Structure.** The model proposed here is a combination of *vicente2020* and *leclere2015*. Its structure is presented in Figure 1. It consists in applying *vicente2020* on inputs prepared following the *leclere2015* approach, as also done in one of the models of Rennie et al. (2011). Instead of using the target and masker signals at the ear, the original speech and noise materials and the BRIRs at the ears are taken as inputs. This allows to re-create the stimuli used in the experiments, except that the early and late targets can be separated. This cannot be done with only the stimuli at the ears. The target BRIR is separated in an early and a late part as in *leclere2015*. To obtain the long-term characteristics of the target speech, an averaged target signal is generated by



**Figure 1.** Structure of the proposed model. The first block represents the signal processing (detailed in Figure 2). The second block represents the *vicente2020* model.



**Figure 2.** Processing of the signals and binaural room impulse responses (BRIRs) in the proposed model.

averaging 60–280 target sentences (depending on the dataset). This averaged target signal is convolved with the early and the late target BRIRs. The early target signal is taken as input in *vicente2020*, as the useful signal. The late target signal is summed with the masker signal, which is the noise material convolved with the masker BRIR. The resulting signal is taken as input in *vicente2020*, as the detrimental signal.

In order to predict the effect of hearing loss, two different methods were tested here to calculate the internal noise. The first method was the one used in *vicente2020*, where the internal noise depends on the listener audiogram and the external sound broadband level. In the second method, this broadband level is replaced by the external sound spectrum, so that the internal noise level in a given frequency band depends on the external sound level in that particular band rather than on its broadband level. This modification influences the predictions only for HI listeners (i.e. here for the datasets from Cueille et al. (2022) and Vicente et al. (2021)). The advantage of the broadband level approach is that it produces an internal noise that has the same spectral shape as the audiograms (Vicente et al., 2020). The spectrum level approach leads to

an internal noise which can have a different spectral shape than the audiograms, but it offers the advantage of better approximating the real sound level in the different frequency bands (which can be quite different from the broadband level).

The processing of the signals and BRIRs is detailed in Figure 2. After convolution with the BRIRs, the target and masker signals need to be calibrated to the presentation level, corresponding to the sound level at which the stimuli were played during the experiment. This level was approximated as the masker level here, because the experiments considered here usually used a masker signal at a fixed level that was higher than the target level. In the proposed model, the calibration of the early and late target signals necessitates that they are multiplied by the gain required to calibrate the level of the “complete” target (the average of the speech sentences convolved with the “complete” target BRIR) to the presentation level.

**Choice of the Early-late Limit.** Leclère et al. (2015) tested several shapes for the temporal windows used to do the early-late separation and showed that they led to very similar results.

Here, only linear windows (as used in the room-independent model *leclere2015*) were considered. Both the ELL and the decay duration can be varied to improve the model performances, so that for a given decay duration an ELL can be found to reach optimal performance (Leclère et al., 2015). Here, it was decided to only vary the ELL and keep the decay duration at a fixed value of 25 ms (the default value used in *leclere2015*), except when the computation method required another decay duration value (see the Kokabi method below). Four methods of ELL computation were tested to determine whether using a room-dependent ELL could improve the predictions, as suggested by Leclère et al. (2015).

First, a room-independent ELL fixed at 30 ms (as in *leclere2015*) was tested. The corresponding model version was called “model-RI” (Room-Independent).

The second approach was based on the physical mixing time. The physical mixing time is defined as the moment when the diffuse reverberation tail begins. In other words it is the moment when the acoustical energy is equidistributed and there is a uniform acoustical energy flux in the room (Lindau et al., 2012). In this approach, the ELL was set to the physical mixing time that was estimated using the method proposed by Abel and Huang (2006), based on the calculation of the echo density profile of each BRIR, meaning that the ELL is room/position-dependent. To calculate the echo density, a rectangular time window of 23 ms slides along the target BRIR. Within each time window, the standard deviation of the sound pressure is calculated. The proportion of samples outside the interval “mean window value  $\pm$  window standard deviation” is calculated. The echo density is defined in each time window as the ratio between this proportion and the proportion of samples which would be outside the window standard deviation if the samples followed a Gaussian distribution. At the beginning of the BRIR, the sound pressure equals zero, so the echo density is also zero. As time increases the sound becomes diffuse and the sound pressure of the BRIR within the time window starts to resemble a Gaussian distribution, so that the echo density tends to one. The physical mixing time/ELL is defined as the instant when the echo density reaches unity. The version of the model using this ELL was called “model-Abel”.

The third model version varied the ELL depending on the BRIR interaural cross-correlation (IACC), based on a linear regression proposed by Kokabi et al. (2018) in a first attempt to develop a room-dependent version of *leclere2015*. They used *leclere2015* on BRIRs simulated by Rennies et al. (2011) (1 room, 4 distances, 2 source positions), as well as BRIRs from a simulated room with varying absorption coefficients (4 absorption coefficients). To predict the SRTs of Rennies et al. (2011) and those measured in their own study, Kokabi et al. (2018) tested *leclere2015* with several ELLs, and did a linear regression in order to find a correlation between the ELLs leading to the best predicted SRTs and the IACCs. The regression  $ELL = 143 - 202 \times IACC$  gave the best results, with a correlation coefficient of 0.74. This ELL

computation was tested in the present study. It was developed with a 1-ms decay duration (Kokabi et al., 2018), and the same decay duration value was thus used here. This version of the model was called “model-Kokabi.”

The fourth method to compute the ELL was based on the BRIR short-term IACC. The similarity of the reflections at the two ears is used to decide whether they are detrimental, assuming that reverberation makes the late target more masking when it is more different at the ears thus limiting binaural unmasking (Lavandier & Culling, 2008; Leclère et al., 2015; Rennies et al., 2019). It is based on the work of Rennies et al. (2019), who showed that reflections might be defined as useful or detrimental depending on whether they are useful/detrimental for SRM and binaural intelligibility, rather than whether they are early or late. Here the IACC was measured in rectangular temporal windows of 23 ms. The ELL was defined as the instant when the short-term IACC first becomes inferior to 0.6. This value was arbitrarily chosen to test the concept, but the estimated ELLs were very similar when it was chosen between 0.7 and 0.5. This version of the model was called “model-IACC.”

## Model Evaluation

The eight versions of the proposed model, corresponding to the four computations of ELL and the two computations of internal noise, were evaluated using the data from seven experiments. The first experiment of Lavandier et al. (2012) was used to for the influence of reverberation on SRM (for stationary noise and NH listeners). The third experiment of Lavandier and Culling (2008) was selected for the influence of reverberation on SRM and the temporal smearing of the target (for stationary noise and NH listeners). The first experiment of Collin and Lavandier (2013) was chosen for the effect of reverberation on dip listening (for frontal sources generating limited binaural effects, and NH listeners). The third experiment of Collin and Lavandier (2013) allowed study the dip listening benefit and the temporal smearing of the target by reverberation (for frontal sources generating limited binaural effects, and NH listeners). The fourth experiment of Collin and Lavandier (2013) was used to combine the effects of dip listening and SRM, even if only at a low reverberation level (for NH listeners). The experiment of Cueille et al. (2022) was chosen to include the effects of reverberation on the temporal smearing of the target and the one of the masker (that affects dip listening) for NH and HI listeners. This experiment used monaural conditions. The proposed model is binaural, but it is also able to make predictions for monaural stimuli. To do so, the model takes as input the signals at one ear, and no signal at the other ear. The dataset 1 of Vicente et al. (2021) allowed to test the influence of reverberation on SRM for stationary and modulated noises, with NH and HI listeners. Among these datasets, the third experiment of Collin and Lavandier (2013) and the experiment of Cueille et al. (2022) cannot be predicted by *leclere2015* and

*vicente2020*, as they include the effect of the temporal smearing of the target speech (that *vicente2020* cannot predict) and the effect of the noise modulations of the masker (that *leclere2015* cannot predict). The experiment from Cueille et al. (2022) also includes the effect of hearing loss, which cannot be predicted by *leclere2015*. The perceptual mechanisms and reverberation effects present in each dataset that justify their use to evaluate the models are summarized in Table 1.

Testing the proposed model requires access to the collected data and the signals used during the experiments, but also to the BRIRs and the original raw speech and noise materials used to create the stimuli. Thus, as a first step, the models were tested here using our own datasets. However, overall, the models were tested on data measured in four different labs, with four different types of speech material, in two languages, using four different procedures to measure speech intelligibility. The experiments from Lavandier et al. (2012) and Lavandier and Culling (2008) were adaptive SRT measurements done at Cardiff University (UK) with unpredictable English sentences from the Harvard Sentence List (IEEE, 1969). The experiments from Collin and Lavandier (2013) were adaptive SRT measurements done at ENTPE (France), with unpredictable French sentences from an open set developed by Raake and Katz (2006). The experiment from Cueille et al. (2022) was done at CRNL (France) and measured full psychometric functions, with French sentences from a closed set matrix-type speech material (Jansen et al., 2012). The dataset from Vicente et al. (2021) consists of adaptive SRT measurements done at Macquarie University (Australia), with English sentences from an open set speech material (Bench et al., 1979).

The predicted SRTs were compared with the measured data, and with the predictions of the original models *leclere2015* and *vicente2020*. The figures comparing the predictions of the proposed model to those of the original models are shown in the paper only for the two datasets that neither *leclere2015* or *vicente2020* can predict: The third experiment of Collin and Lavandier (2013) and the experiment of Cueille et al. (2022). For the other datasets the corresponding figures can be found in the Supplemental Materials (see Supplemental Materials at [URL will be inserted by SAGE]).

The four ELL calculation methods were tested for all datasets, except the one from Cueille et al. (2022). The Kokabi and IACC versions could not be tested with this dataset because the stimuli are monaural. The two methods of internal noise calculation were only compared with the two datasets involving HI listeners, as the internal noise does not affect the predictions for NH listeners. For each dataset and for all model versions, the Pearson's correlation coefficient between the average measured and predicted SRTs was calculated, along with the mean and largest absolute errors between average data and predictions. Furthermore, the individual predictions of the HI listeners are shown in the Supplemental Materials for the two datasets involving HI

listeners, along with the Pearson correlation coefficient between the individual measured and predicted SRTs and the mean and largest absolute errors between individual data and prediction across conditions. The individual predictions of the NH listeners are not shown, as the models give very similar predictions for the NH listeners, because they do not use individual inputs other than the audiograms. Thus, they cannot predict the individual variability of the NH listeners (Lavandier et al., 2021).

### Implementation of the Predictions

Depending on the dataset the target signal was created by averaging between 60 and 280 sentence waveforms. The datasets involving only NH listeners did not give information on the presentation level used during data collection. The presentation level was fixed at 65 dB SPL for these datasets. This did not affect the predictions as the presentation level only affects the model through the internal noise that has an effect only for the predictions of HI listeners (unless very soft levels are considered, which was not the case here). When the audiograms of the NH listeners were unavailable, their hearing thresholds were fixed at 0 dB HL at all frequencies.

To compare absolute SRTs and binaural ratios (the model output), a reference must be chosen. This reference does not influence the relative differences in predictions across conditions. The average measured SRT within each dataset was chosen as reference (Lavandier et al., 2022). Before comparison to the measured data, the binaural ratios were inverted (higher ratios reflect better intelligibility and lower SRTs). Inverted binaural ratios were offset by subtracting their mean and adding the corresponding average SRT across conditions and listeners (the reference), thus aligning the inverted binaural ratios with the measured SRTs. The predictions of the models *leclere2015* and *vicente2021* were compared to those of the proposed model in order to verify its backward compatibility, as well as to illustrate the improvements between the previous and new models. The models *leclere2015* and *vicente2020*, implemented and made available by Lavandier et al. (2022), were used here using the same assumptions as those of the proposed model (whenever relevant).

### Datasets used to Test the Models and Results of the Predictions

The datasets included only NH listeners, unless otherwise specified. The proposed model versions based on the external broadband level and spectrum level give identical predictions for NH listener groups, so that four new model versions are compared. For the datasets with a HI group, the corresponding difference in internal noise implementation leads to eight tested new model versions. The performance statistics for

**Table 1.** Summary of the Perceptual Mechanisms and Reverberation Effects Present in Each Tested Dataset. X Indicates the Presence of a Mechanism/effect.

Dataset	Hearing loss	Reverb-target speech	Dip listening	Reverb-dip listening	Spatial release from masking (SRM)	Reverb-SRM
Lavandier et al. (2012)					X	X
Lavandier and Culling (2008)		X			X	X
Collin and Lavandier (2013) Exp.1			X	X		
Collin and Lavandier (2013) Exp.3		X	X	X		
Collin and Lavandier (2013) Exp.4			X		X	
Cueille et al. (2022)	X	X	X	X		
Vicente et al. (2021)	X		X	X	X	X

the different model versions are shown in Table 2. The estimated ELL obtained with the four ELL separation methods can be found in Table 3.

### Lavandier et al. (2012)

In experiment 1 of Lavandier et al. (2012), the stimuli were convolved with BRIRs measured in a real room and spectral envelope impulse responses (SEIRs). The SEIRs were obtained by removing the temporal characteristics of the BRIRs, in order to remove the ITDs while preserving the ILDs, so that better-ear listening is possible while binaural unmasking is not. A single stationary noise was tested at two distances (0.65 m, “near”, and 5 m, “far”) and three orientations ( $-25^\circ$ , “left”,  $0^\circ$ , “front”,  $+25^\circ$ , “right”). The target speech was always at the same position (near right). The measured SRTs and predictions of the four versions of the proposed model can be found in Figure 3. Including this dataset allows to test the model predictions for the effect of SRM, which is highlighted by comparing the SRTs in the left and front conditions to those in the co-located condition (near right). An interaction between reverberation and SRM was also observed, as there was less SRM in the far conditions, which allows to test the models for the effect of reverberation on SRM.

As expected, *vicente2020* and *leclere2015* accurately predict this dataset (Supplemental Figure 1) and give performance statistics similar to the proposed model (Table 2).

The four versions of the proposed model give similarly accurate predictions. The Kokabi and IACC versions overestimate the SRTs by 0.5 dB in the BRIRs conditions and underestimate them by 0.5 dB in the SEIRs conditions. The performance statistics in Table 2 do not reflect these small differences and show similar accuracy for the four versions. It is interesting to note that even though their predictions are similar, the four model versions estimate different ELLs for this dataset. As can be seen in the Table 3, the Abel version considered the whole target BRIR as useful (as the echo density

never reached 1), while the Kokabi and IACC versions estimated an ELL lower than the room-independent ELL at 30 ms.

### Lavandier and Culling (2008)

In experiment 3 of Lavandier and Culling (2008), a single stationary noise was always spatially separated from the target speech, the noise being  $65^\circ$  to the left of the listener and the target being  $65^\circ$  to its right. The room absorption coefficients used for the target and noise were controlled separately for each source, testing four coefficients: 1, 0.7, 0.5, 0.2. The measured SRTs and model predictions are shown in Figure 4. The inclusion of this dataset allows to test the model for the effect of the temporal smearing of the target speech, as well as the influence of reverberation on SRM. At low target absorption coefficients, the SRT increases as the target absorption coefficient decreases (reverberation increases), highlighting the temporal smearing of the target speech. When the masker absorption coefficient decreases, the SRT increases, highlighting the detrimental effect of reverberation on SRM.

The *leclere2015* model has similar performance statistics as the proposed model (Table 2). On the other hand, *vicente2020* is inaccurate (Supplemental Figure 2). This is expected, as *vicente2020* does not consider the early and late targets separately and cannot predict the effect of the temporal smearing of the target speech.

The room-independent and Abel model versions give results similar to those of *leclere2015*. They overestimate the SRT when the target absorption coefficient is 1 and slightly underestimate it when it is 0.5. The Kokabi version gives similar predictions, but is even less accurate when the target absorption coefficient is 0.5. The IACC version gives the least accurate predictions. This model version underestimates the SRT when the target absorption coefficient is 1 and 0.7, and largely overestimates it when it is 0.2. The performance statistics also show this difference in performance, with the Kokabi

**Table 2.** Pearson Correlation Coefficient, Mean and Largest Errors Obtained with the Two Original Models (*vicente2020* and *leclere2015*), as well as the Four Versions of the Proposed Model with the Internal Noise Depending on the External Sound Spectrum.

Dataset	Vicente2020	Leclere2015	Model-RI	Model-Abel	Model-Kokabi	Model-IACC
Lavandier et al.(2012)	0.95	0.97	0.98	0.98	0.97	0.98
	0.4 dB	0.3 dB	0.2 dB	0.2 dB	0.3 dB	0.2 dB
	0.6 dB	0.9 dB	0.6 dB	0.6 dB	0.8 dB	0.7 dB
Lavandier and Culling (2008)	0.48	0.86	0.85	0.87	0.8	0.85
	1.2 dB	0.8 dB	0.8 dB	0.8 dB	1 dB	1.8 dB
	2.3 dB	1.3 dB	1.4 dB	1.3 dB	1.9 dB	4.2 dB
Collin and Lavandier (2013) Exp.1	0.85	0.49	0.87	0.87	0.89	0.87
	0.5 dB	0.6 dB	0.4 dB	0.4 dB	0.3 dB	0.4 dB
	1.3 dB	1.5 dB	1.2 dB	1.2 dB	0.9 dB	1.1 dB
Collin and Lavandier (2013) Exp.3	<b>0.57</b>	<b>0.98</b>	<b>0.99</b>	0.99	0.85	0.99
	<b>3.6 dB</b>	<b>2.2 dB</b>	<b>1.2 dB</b>	0.3 dB	3.1 dB	1 dB
	<b>4.5 dB</b>	<b>3.2 dB</b>	<b>1.4 dB</b>	0.7 dB	3.6 dB	1.2 dB
Collin and Lavandier (2013) Exp.4	0.93	0.87	0.91	0.91	0.93	0.91
	0.6 dB	0.7 dB	0.6 dB	0.6 dB	0.4 dB	0.5 dB
	1.4 dB	1.8 dB	1.4 dB	1.5 dB	0.9 dB	1.4 dB
Cueille et al. (2022)	<b>0.59</b>	<b>0.73</b>	<b>0.93/0.95</b>	0.87/0.89	N/A	N/A
	<b>2.4 dB</b>	<b>2 dB</b>	<b>0.9 dB/1.4 dB</b>	1.2 dB/1.5 dB		
	<b>6.1 dB</b>	<b>3.7 dB</b>	<b>2.4 dB/3 dB</b>	3 dB/3.6 dB		
Vicente et al. (2021)	0.92	0.39	0.87/0.93	0.88/0.93	0.87/0.92	0.87/0.93
	0.9 dB	1.7 dB	1.1 dB/0.9 dB	1.1 dB/0.9 dB	1.1 dB/0.9 dB	1.1 dB/0.9 dB
	2 dB	4.5 dB	2.8 dB/1.9 dB	2.7 dB/1.9 dB	2.7 dB/2.5 dB	2.7 dB/1.9 dB

Note. The performances of the model versions with the internal noise depending on the external sound broadband level are presented in italic, only for the two last datasets for which this change affects the predictions. The performances in bold highlight the two datasets that the original models do not predict well.

**Table 3.** Estimated Early-late Limits (in ms) of the Three Room-dependent Model Versions for the Seven Tested Datasets.

Target BRIR	Abel	Kokabi	IACC
LAV12:BRIR	37	6	11
LAV12:SEIR	42(end)	0	11
LAV08:1	243(end)	0	12
LAV08:07	243(end)	12	11
LAV08:05	26	49	0
LAV08:02	21	26	0
COL13:exp1	23	0	11
COL13:exp3-065	25	22	11
COL13:exp3-1000	14	104	0
COL13:exp4	32	0	11
CUE22:0.15	65	N/A	N/A
CUE22:0.5	70	N/A	N/A
CUE22:0.8	75	N/A	N/A
CUE22:1.1	39	N/A	N/A
CUE22:1.5	50	N/A	N/A
VIC21:+20	243(end)	0	5
VIC21:+40	243(end)	0	5
VIC21:-90	243(end)	0	5

Note. Each line corresponds to a target BRIR. Datasets are separated by double lines. BRIR = binaural room impulse response; IACC = interaural cross-correlation; SEIR = spectral envelope impulse response.

and IACC versions leading to larger mean and largest prediction errors. For this dataset, the Abel version considered the whole target BRIR as useful for the two highest target absorption coefficients, and estimated ELLs close to 30 ms for the two lowest coefficients (Table 3). The Kokabi version followed an

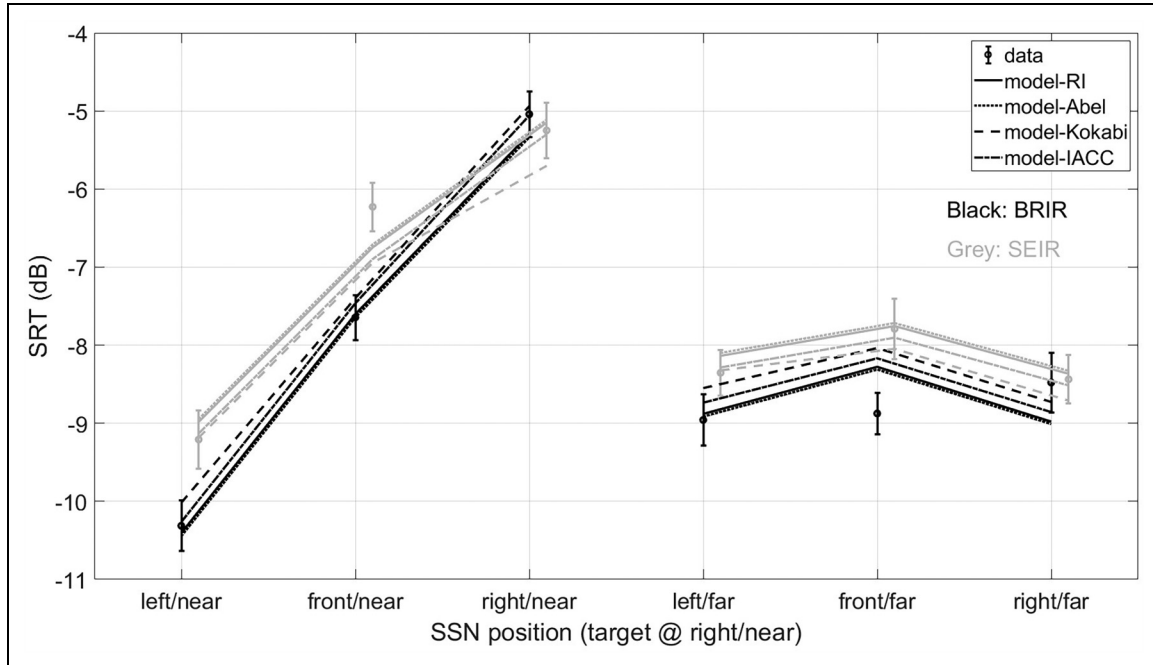
inverse trend, estimating low ELLs (below 15 ms) for the highest absorption coefficient and larger ELLs (49 and 26 ms) for the two lowest absorption coefficients. The IACC version estimated low ELLs (below 15 ms) for all conditions.

### Collin and Lavandier (2013) - Experiment 1

In experiment 1 of Collin and Lavandier (2013), SRTs were measured with stationary noise or noise modulated by the envelope of a speech signal comprised of one, two or four voices. The interferer was tested at three distances (65, 125, and 500 cm) in front of the listener in a lecture hall. Increasing its distance increased the relative amount of reverberation for the interferer. The target was always at 65 cm in front of the listener. The measured SRTs and model predictions can be found in Figure 5. This dataset tests the models' ability to predict the dip listening benefit, as well as the effect of reverberation on this dip listening benefit. A dip listening benefit is observed, as the SRTs are lower when the masker is modulated. This dip listening benefit is reduced by reverberation when the interferer distance increases. Note that room coloration also made the noises less masking when their distance increased, counterbalancing the reduction of dip listening for the modulated noises (for more discussion on this effect, see Collin and Lavandier (2013)).

The *vicente2020* model predicts this dataset as accurately as the proposed model, while *leclere2015* gives inaccurate predictions (Supplemental Figure 3) with a correlation





**Figure 3.** Mean SRTs with standard errors across listeners measured by Lavandier et al (2012). Predicted SRTs are displayed for the proposed model with four ELL calculation methods. A small horizontal shift was added to help differentiate the conditions involving BRIRs and SEIRs. Note. SRT = speech reception threshold; ELL = early-late limit; BRIR = binaural room impulse response; SEIR = spectral envelope impulse response.

between data and prediction of 0.49. This was expected as *leclerc2015* cannot predict the effect of masker modulations.

The four versions of the proposed model give accurate predictions for this dataset. The room-independent, Abel and IACC versions underestimate the SRT by 0.5 dB in the conditions with one-voice and two-voice modulated maskers, and overestimate it by 1 dB in the conditions with a stationary masker. In other words, the dip listening benefit appears slightly overestimated. The Kokabi version is a bit more accurate in this regard. For this dataset, the Abel version estimated an ELL of 23, against 0 ms for the Kokabi version (in other words, the early target BRIR did not have a constant portion and consisted only of the decaying part of 1 ms), and 11 ms for the IACC version (Table 3).

### Collin and Lavandier (2013) - Experiment 3

In experiment 3 of Collin and Lavandier (2013), a stationary noise and a one-voice modulated noise were tested, always close to the listener (at 65 cm). The target was tested at two distances: 65 cm and 10 m. The target and the interferer were always in front of the listener in a L-shaped room. The measured SRTs and model predictions are shown in Figure 6. This dataset is included to test the models for predicting the dip listening benefit and the temporal smearing of the target speech simultaneously. The temporal smearing of the target speech is highlighted by the SRT increase between the 65-cm and 10-m conditions. The dip listening benefit is highlighted by the SRT difference between

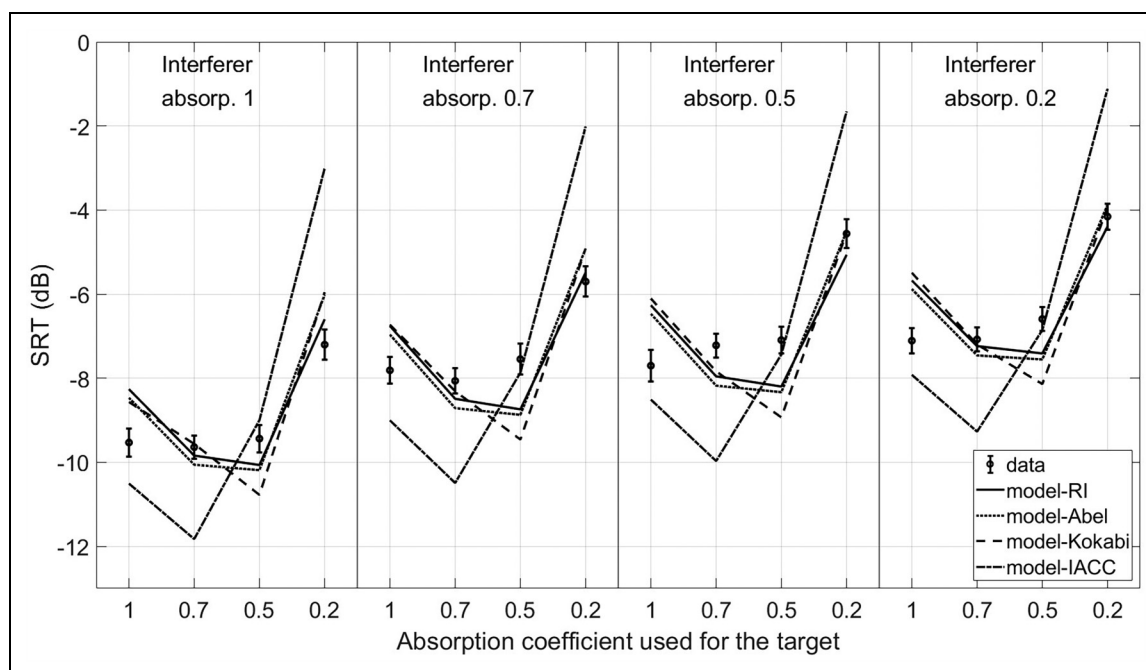
the stationary and one-voice modulated masker. There was no significant interaction between the target distance and interferer modulations.

The Abel model version gives the best predictions, but appears to underestimate the SRTs by 0.5 dB at the larger distance (10 m), underestimating the effect of the temporal smearing of the target speech. This tendency was found at a larger degree with the room-independent and Kokabi versions, the Kokabi version being the worst. The performance statistics show that these two versions lead to larger prediction errors. The IACC version follows an inverse trend, overestimating the temporal smearing of the target speech. For this dataset, the Abel version estimated ELLs of 25 and 14 ms for the 65 cm and 10 m conditions, respectively. The Kokabi version estimated ELLs of 22 and 104 ms. The IACC version estimated ELLs of 11 and 0 ms.

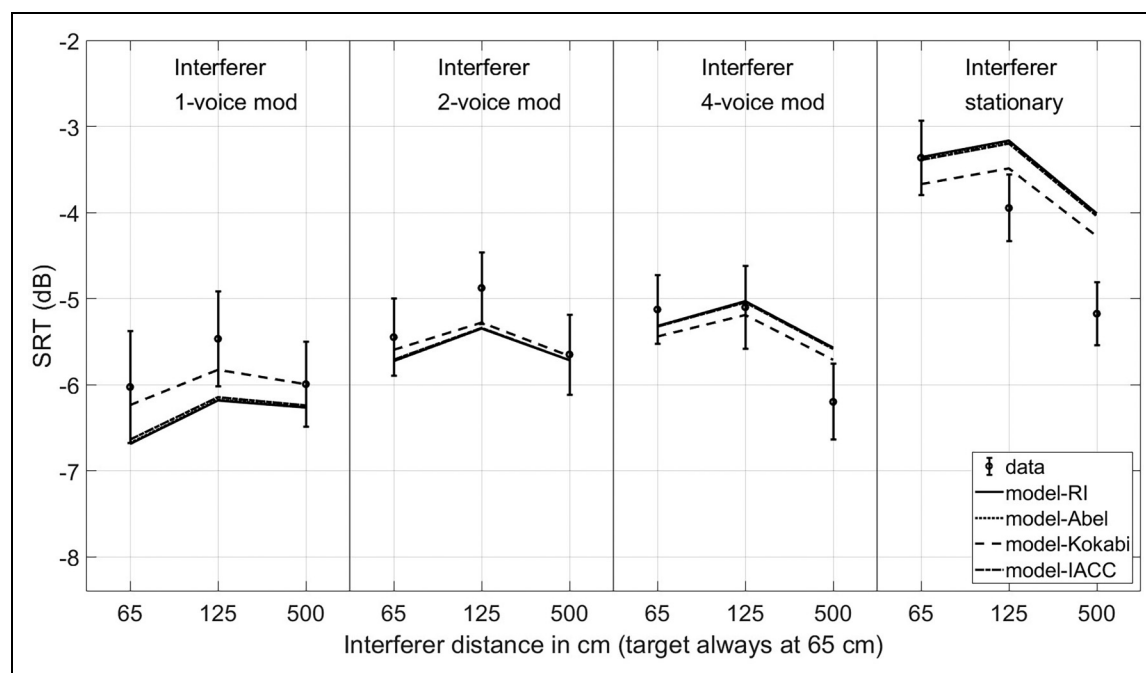
Contrary to the new model, *vicente2020* and *leclerc2015* give inaccurate predictions for this dataset, as can be seen in Figure 7. The *vicente2020* model leads to a correlation between data and prediction of only 0.59 and large prediction errors (Table 2), because it cannot account for the effect of the temporal smearing of the target speech. The *leclerc2015* model provides a correlation of 0.98, but also large prediction errors, because it cannot account for the effect of masker modulations and dip listening.

### Collin and Lavandier (2013) - Experiment 4

In experiment 4 of Collin and Lavandier (2013), a stationary noise, a one-voice modulated noise and a two-voice



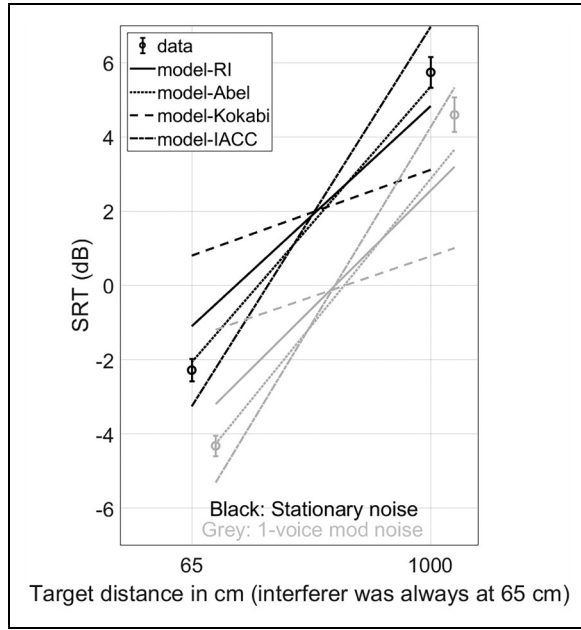
**Figure 4.** Mean SRTs with standard errors across listeners measured by Lavandier and Culling (2008). Predicted SRTs are displayed for the proposed model with four ELL calculation methods. *Note.* SRT = speech reception threshold; ELL = early-late limit.



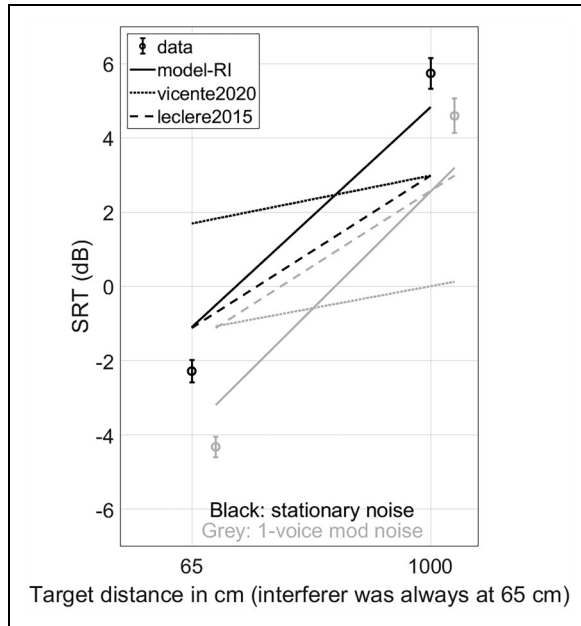
**Figure 5.** Mean SRTs with standard errors across listeners measured by Collin and Lavandier (2013) in their first experiment. Predicted SRTs are displayed for the proposed model with four ELL calculation methods. *Note.* SRT = speech reception threshold; ELL = early-late limit.

modulated noise were tested. The target was always in front of the listener. The stationary and two-voice modulated noises were tested in three configurations: In front (co-located with

the target), asymmetrical (+25°) and symmetrical (+25° and -25°). The one-voice modulated noise was only tested in the co-located and asymmetrical configurations. All sources



**Figure 6.** Mean SRTs with standard errors across listeners measured by Collin and Lavandier (2013) in their third experiment. Predicted SRTs are displayed for the proposed model with four ELL calculation methods. A small horizontal shift was added to help differentiate the conditions involving stationary and modulated noise. Note. SRT = speech reception threshold; ELL = early-late limit.



**Figure 7.** Same as Figure 6 but with the predictions of *vicente2020* and *leclere2015* (as well as *model-RI* for comparison). The predictions of *leclere2015* are identical with stationary and modulated noise, as the noise modulations are not taken as input (only the artificial horizontal offset differentiate them here).

were tested at the same fixed distance (65 cm) in a meeting room (the room also used in experiment 1 of Lavandier et al. (2012)). The measured SRTs and model predictions can be found in Figure 8. This dataset is used to combine the effects of dip listening and SRM, even if only at a low reverberation level. The dip listening benefit can be seen with the lower SRTs in the conditions with modulated noises. The SRM is indicated by the lower SRTs in the asymmetrical configuration compared to the co-located condition. The SRM was much reduced in the symmetrical configuration.

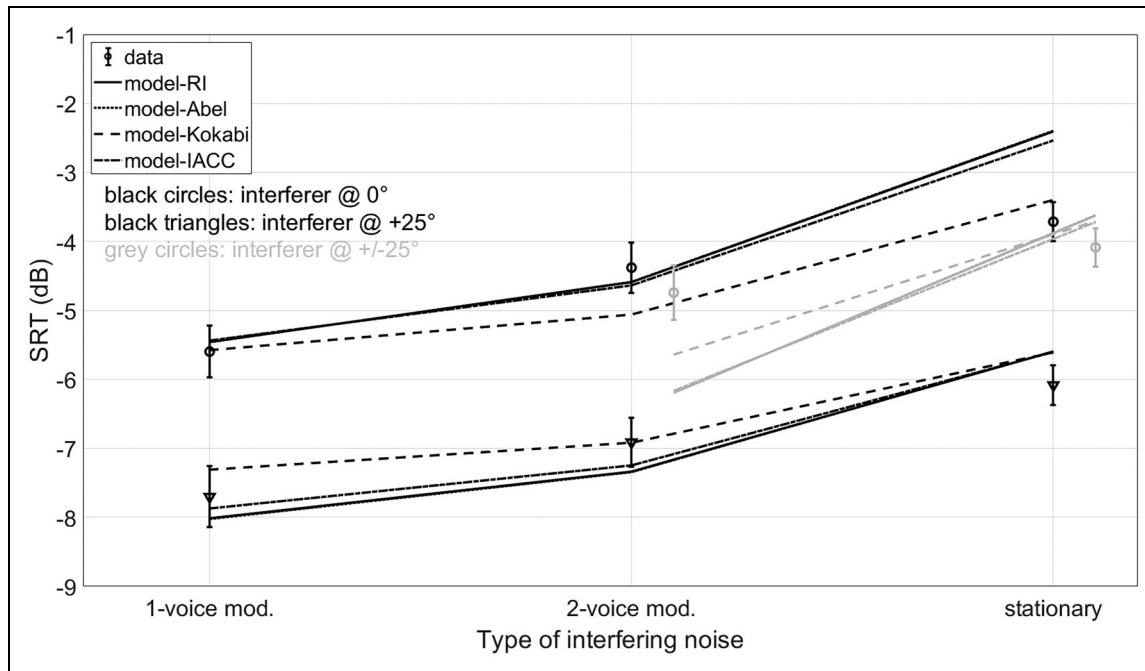
The *vicente2020* model accurately predicts this dataset (Table 2), with performance statistics similar to those of the proposed model. The *leclere2015* model is able to predict the effect of spatial configuration, but not the effect of masker modulations (Supplemental Figure 4). Its performance statistics are slightly worse than those of the other models.

The room-independent, Abel and IACC versions show accurate predictions. They overestimate the SRTs by 1 dB for the stationary noise, implying that the dip listening benefit is slightly overestimated. The SRTs are also underestimated for the two-voice-modulated noise in the symmetrical configuration ( $+/- 25^\circ$ ). The Kokabi version might provide more accurate predictions, as also reflected by the performance statistics. The ELLs estimated for this dataset were 32, 0 and 11 ms, for the Abel, Kokabi and IACC versions, respectively.

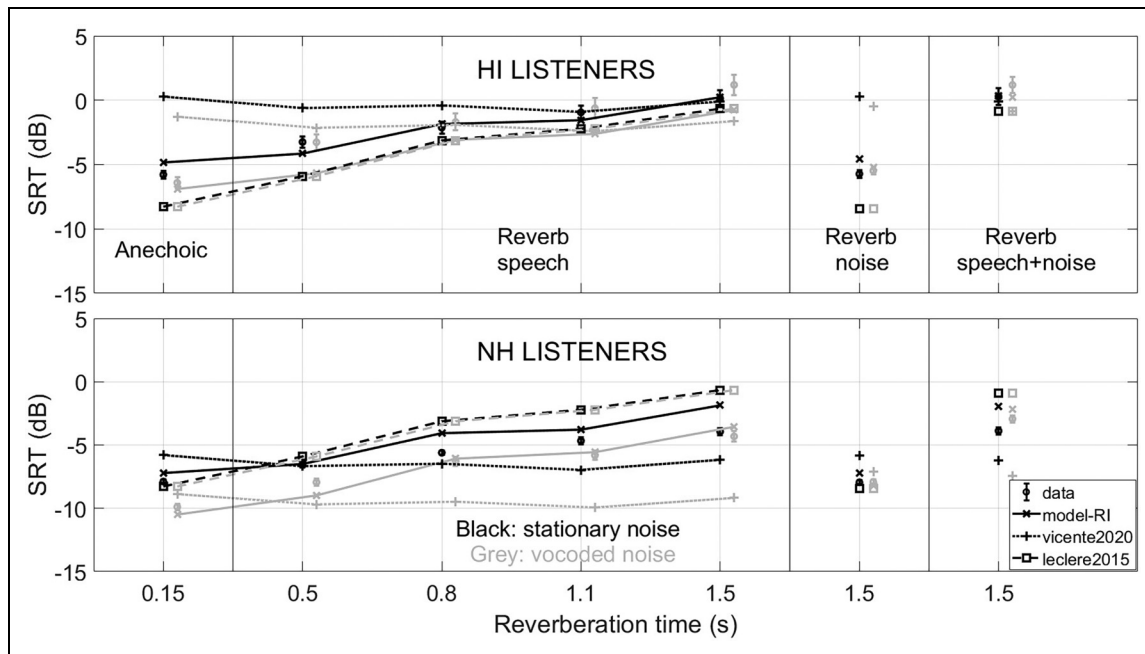
### Cueille et al. (2022)

Cueille et al. (2022) tested 32 NH and 32 HI listeners. The HI listeners had a mean four-frequency pure tone average (PTA4) of  $45.2 \pm 11.3$  dB HL. All the stimuli were monaural. A stationary noise and an unintelligible two-voice vocoded speech (as a proxy for modulated noise) were tested. Reverberation was applied either only on the target speech (four reverberation times tested), only on the interfering noise (one reverberation time), or on both target and noise (one reverberation time). An anechoic condition was also tested. The stimuli for the HI listeners were amplified using a linear frequency-dependent gain set with the NAL-RP prescription rule (Dillon, 2012). The comparison between the two original models and the room-independent version of the proposed model using the external sound spectrum as input is presented in Figure 9. The measured SRTs and the predictions from the four ELL versions of the proposed model with the external spectrum as input are shown in Figure 10, while those using the external broadband level appear in Figure 11.

This dataset was chosen to include the effects of reverberation on the temporal smearing of the target speech as well as the effect of reverberation on the dip listening benefit, for NH and HI listeners. In Figures 9 to 11, the effect of hearing loss can be seen with the lower SRTs of the NH listeners. The effect of the temporal smearing of the speech is highlighted by the SRT increases when the reverberation time used for the target speech increases. Very limited dip listening benefit (difference between black and grey SRTs) was observed in the data, only for the NH



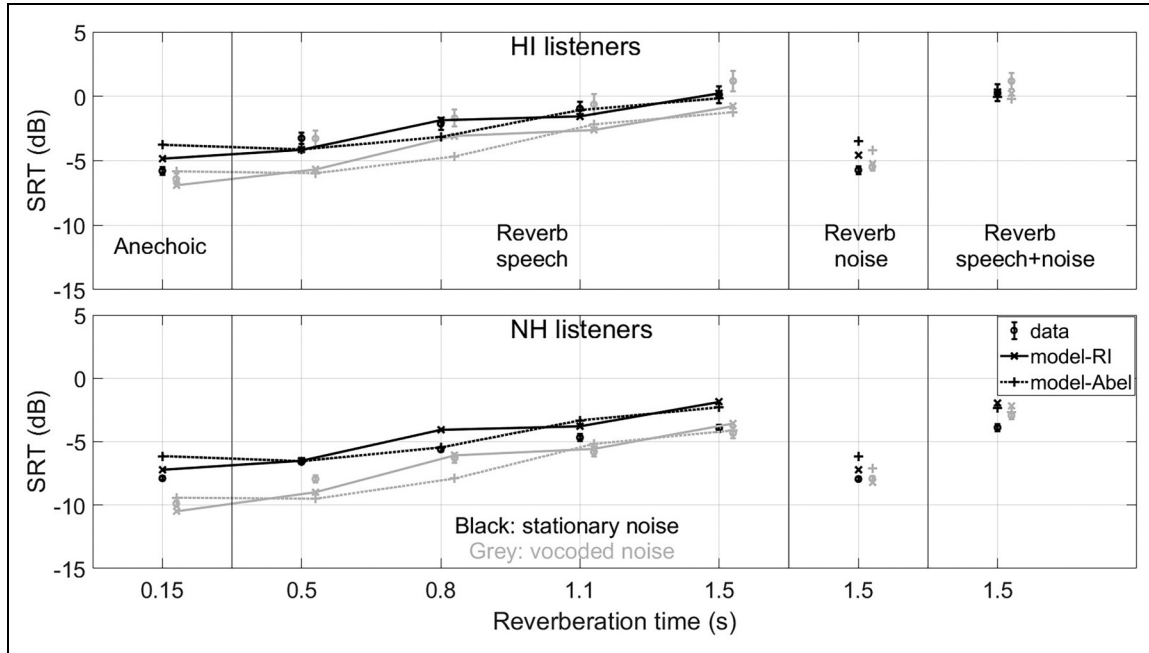
**Figure 8.** Mean SRTs with standard errors across listeners measured by Collin and Lavandier (2013) in their fourth experiment. Predicted SRTs are displayed for the proposed model with four ELL calculation methods. A small horizontal shift was added to help differentiate the conditions involving symmetric maskers. Note. SRT = speech reception threshold; ELL = early-late limit.



**Figure 9.** Mean speech reception thresholds (SRTs) with standard errors across listeners measured by Cueille et al. (2022). Predicted SRTs are displayed for *vicente2020* and *leclere2015*, as well as model-RI (using the external sound spectrum) for comparison. The predictions of *leclere2015* are identical with stationary and modulated noise. A small horizontal shift was added to help differentiate the conditions involving stationary and modulated noise.

listeners in the less reverberant conditions. Cueille et al. (2022) found a larger dip listening benefit at lower SNRs, considering the full psychometric functions of the listeners.

Figure 9 shows that *vicente2020* and *leclere2015* inaccurately describe this dataset, while the room-independent version of the proposed model gives accurate predictions. The



**Figure 10.** Mean SRTs with standard errors across listeners measured by Cueille et al. (2022). Predicted SRTs are displayed for the proposed model with the internal noise depending on the external sound spectrum, for two ELL calculation methods (the two binaural methods could not be tested with monaural stimuli). Note. SRT = speech reception threshold; ELL = early-late limit.

*vicente2020* model leads to a correlation of 0.59 and large mean and largest prediction errors, because it does not account for the temporal smearing of the target speech. The *leclerc2015* model provides a correlation of 0.73 and large prediction errors, because it cannot predict the effect of hearing loss and the large differences between the SRTs of the NH and HI listeners.

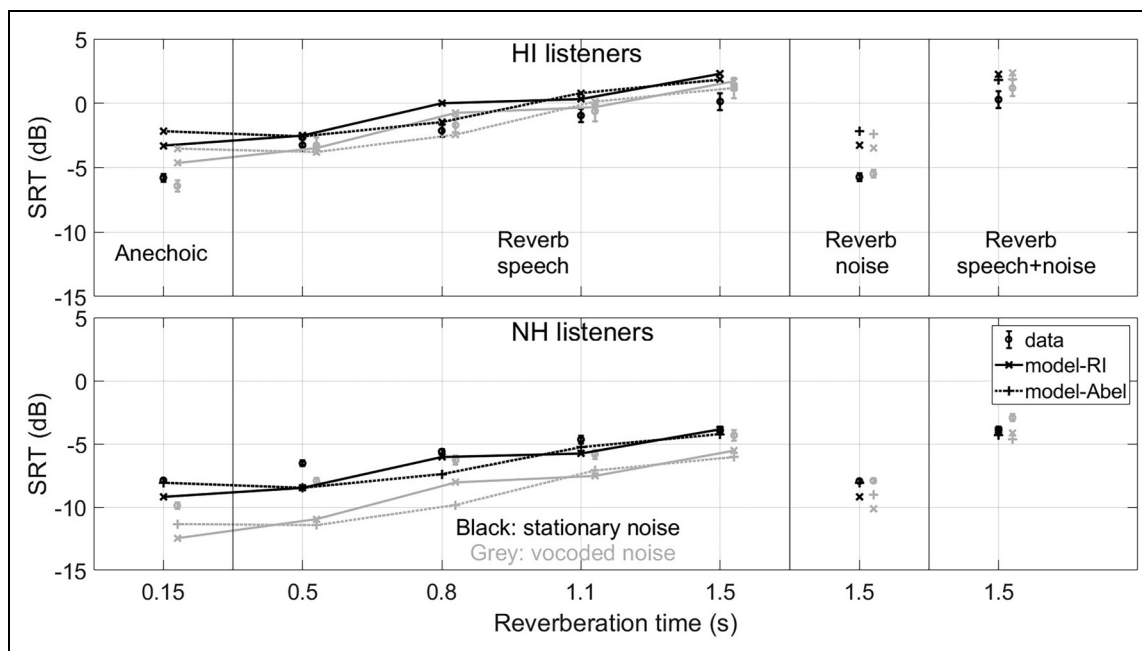
The Kokabi and IACC model versions could not be tested here because the stimuli are monaural. The room-independent model version with an internal noise based on the external sound spectrum can predict all the trends of the dataset, even if it overestimates the dip listening benefit by 1 dB (Figure 10). The Abel version gives similar predictions, and underestimates the SRT variations between the conditions with low and high reverberation times, thus leading to slightly lower correlation between data and predictions and a larger maximum prediction error. For this dataset, the Abel version estimated large ELLs (between 39 and 75 ms). The larger ELLs were obtained for the intermediate reverberation times (0.5 and 0.8 s), while the smaller ELLs were obtained for the highest reverberation times (1.1 and 1.5 s).

Figure 11 shows that the model versions that use the external sound broadband level were less accurate than the versions using the external sound spectrum. When focusing on the room-independent versions, taking the broadband level as input led to an overestimation of the SRT difference between the NH and HI listeners by 3 dB (Figure 11), while the version using the external spectrum underestimated this difference by 1 dB (Figure 10). This is also apparent in the

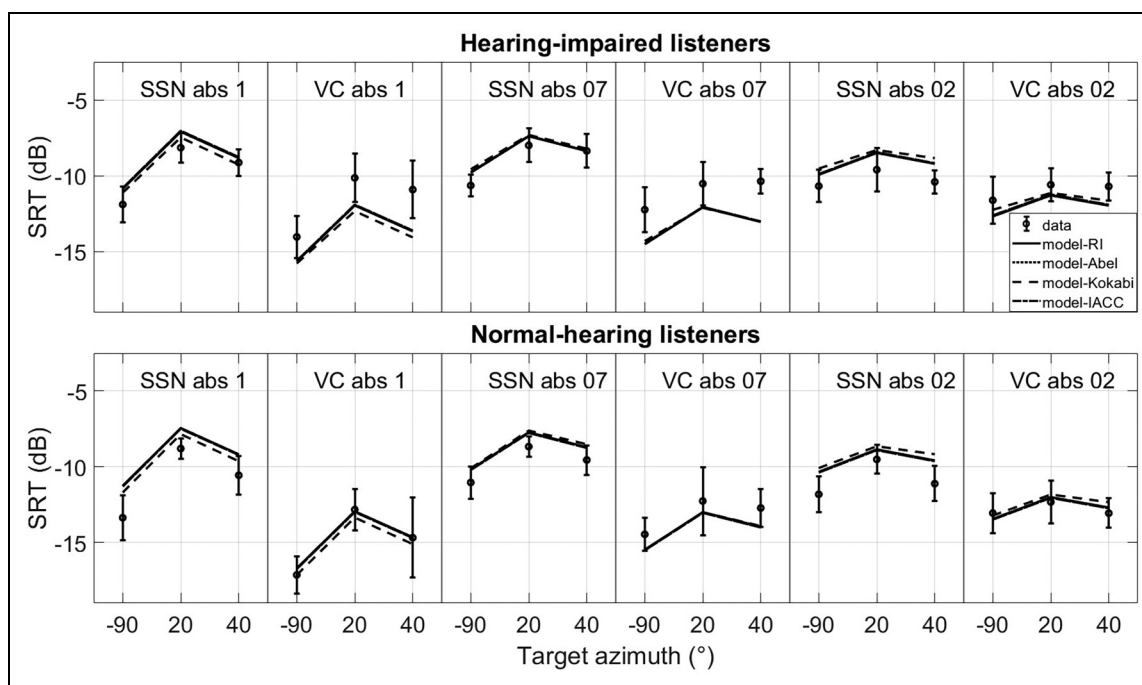
individual predictions (Supplemental Figure 5), where the SRTs of some listeners were largely overestimated in all conditions when using the external broadband level, but not when using the external spectrum (correlation of 0.73, mean error of 3.9 dB, largest error of 16.2 dB with the broadband level; correlation of 0.75, mean error of 2.8 dB and largest error of 9.6 dB with the spectrum).

### Vicente et al. (2021)

Vicente et al. (2021) tested 12 NH and 8 HI listeners. The HI listeners had a PTA4 of  $30.6 \pm 5.8$  dB HL. In the dataset 1 of their experiment, they used stationary speech-shaped noise and noise-vocoded speech as maskers. Three levels of reverberation were used for the masker, with room absorption coefficients of 1, 0.7, and 0.2. The target sentences were anechoic. The masker remained at 6.16 m of the listener, with an azimuth of  $16.4^\circ$ . The target sentences were simulated at 2 m of the listeners, with three tested azimuths:  $-90^\circ$ ,  $20^\circ$  and  $40^\circ$ . The stimuli of the HI listeners were amplified with the NAL-RP prescription rule. The measured SRTs and predictions from the four ELL versions of the model with the external spectrum as input are shown in Figure 12, while Figure 13 compares the two room-independent versions. This dataset was used for the influence of reverberation on SRM for stationary and modulated noises, with NH and HI listeners. The effect of hearing loss is highlighted with the higher SRTs of the HI listeners. The dip listening benefit is highlighted with lower SRTs for the noise-vocoded speech than



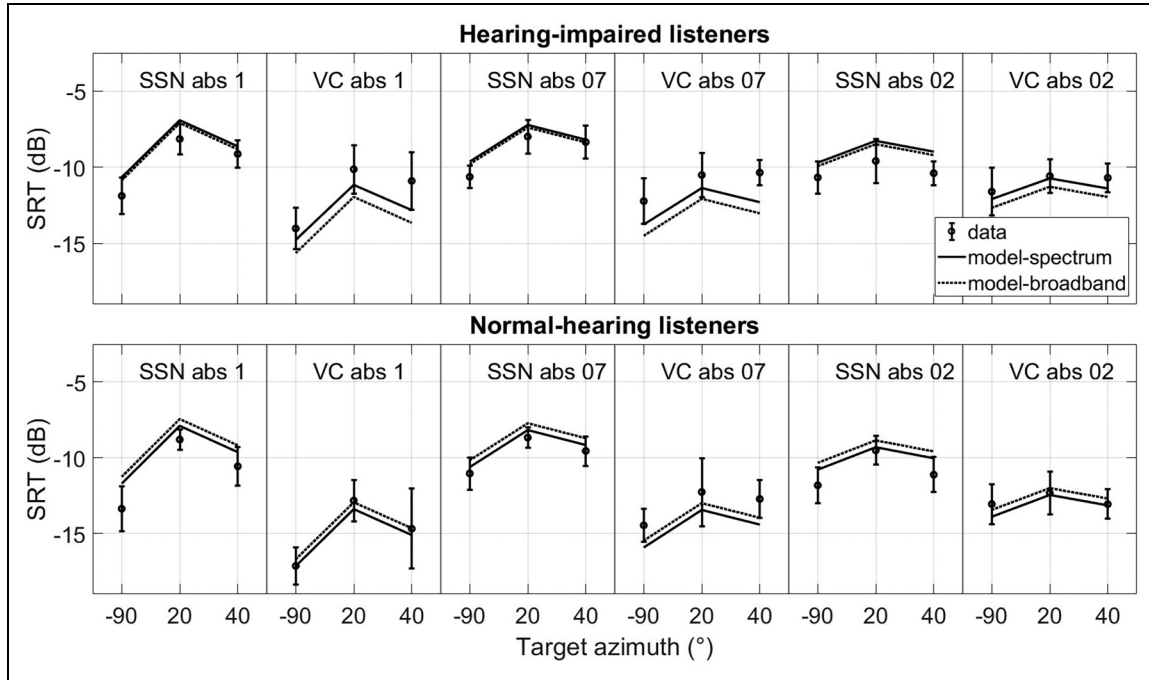
**Figure 11.** Same as Figure 10 but for the model versions with the internal noise depending on the external sound broadband level.



**Figure 12.** Mean SRTs with standard errors across listeners measured by Vicente et al. (2021) using a stationary noise (SSN) or VS masker. Predicted SRTs are displayed for the proposed model with the internal noise depending on the external sound spectrum, for four ELL calculation methods. These methods give very similar predictions, so that the corresponding lines mostly overlap. Note. SRT = speech reception threshold; ELL = early-late limit; VS = vocoded-speech.

for the speech-shaped noise. The SRM is highlighted with lower SRTs for the target azimuths  $-90^\circ$  and  $40^\circ$ . Increasing reverberation reduced both SRM and dip listening.

The *vicente2020* model accurately predicts this dataset, with predictions very similar to the versions of the proposed model using the external sound broadband level (Supplemental



**Figure 13.** Same as Figure 12 but for the predictions of the RI-model versions based on the external spectrum and broadband level.

Figure 6). This was expected, as the two models are equivalent when the target speech is not temporally smeared by reverberation. The *leclerc2015* model gives poor performances, as it does not take into account the effects of hearing loss and noise modulations.

The four ELL separation methods give very similar predictions both for the model versions based on the external sound spectrum (Figure 12) and those based on the external broadband level (not shown). The Kokabi method has slightly worse performances than the other three methods, underestimating the SRTs at low reverberation levels. All model versions seem to overestimate the dip listening benefit by about 2 dB in the condition where the room absorption coefficient is equal to 1. For this dataset, the Abel version considered the whole target BRIR as useful, while the Kokabi and IACC versions estimated ELLs of 0 and 5 ms, respectively.

Figure 13 shows that the two internal noise implementations give very similar predictions, the version based on the external spectrum being slightly less accurate. Using the external spectrum led to an underestimation of the SRT difference between the NH and HI listeners by 1 dB, while using the broadband level gave accurate predictions. The two versions gave similar performance statistics for the individual predictions of the HI listeners (Supplemental Figure 7; correlation of 0.71, mean error of 1.5 dB, largest error of 5.3 dB with the broadband level; correlation of 0.70, mean error of 1.5 dB and largest error of 5.8 dB with the spectrum).

## Discussion

### Comparison with the Original Models

The *vicente2020* model was inaccurate for the experiments of Lavandier and Culling (2008), Cueille et al. (2022), and the third experiment of Collin and Lavandier (2013). In these experiments the speech was temporally smeared, either by increasing the room absorption coefficients used for the speech (Cueille et al., 2022; Lavandier & Culling, 2008) or by increasing the distance between the speech source and the listener (Collin & Lavandier, 2013). The *leclerc2015* model gave inaccurate results for the datasets with modulated noise or HI listeners. This concerns the first and third experiments of Collin and Lavandier (2013), which tested NH listeners in conditions with voice-modulated noise, as well as the experiments of Cueille et al. (2022) and Vicente et al. (2021), which tested HI listeners with vocoded speech. These results were expected, as *vicente2020* does not discriminate between early and late target reflections, and *leclerc2015* does not consider the masker modulation characteristics as input nor the hearing profile of the listener. In contrast, the room-independent version of the proposed model gave predictions as accurate as *vicente2020* for the datasets with a modulated interferer or HI listeners, and as accurate as *leclerc2015* for the datasets with a temporally smeared target speech. It was also able to describe the data from Cueille et al. (2022) and experiment 3 of Collin and Lavandier (2013), something that neither *vicente2020* nor *leclerc2015* can do. The proposed model thus combines the advantages

of each original model by accounting for modulations in the noise and the listener audiogram (as the *vicente2020* model), as well as the temporal smearing of the speech by reverberation (as the *leclere2015* model). In doing so, it extends their performance and generalizability to a larger range of conditions.

### Comparison of the Model Versions

The two model variations discussed here are the target BRIR separation method and the internal noise calculation. The former affects the prediction of the effect of the temporal smearing of the target speech, while the latter affects the prediction of the effect of hearing loss.

Changes in ELL affected the predictions for BRIRs with strong reverberation, but usually had little to no effect for BRIRs with little reverberation. As BRIRs with little reverberation have little to no reflections, variations of ELL will not affect the predictions as long as the direct sound remains in the early window of the BRIR. This can be seen with the datasets from Lavandier et al. (2012) and Vicente et al. (2021) where large changes in ELL between model versions had almost no effect on the predictions.

The room-independent and Abel versions gave similar results for the datasets from Lavandier et al. (2012), Lavandier and Culling (2008), and the first and fourth experiments of Collin and Lavandier (2013). They differed only for two datasets. For the third experiment of Collin and Lavandier (2013), the Abel version was more accurate, while the room-independent version underestimated the effect of reverberation on the target. For the experiment of Cueille et al. (2022), the room-independent version was more accurate, with the Abel version underestimating the effect of reverberation on the target. For this dataset the ELL estimated with the Abel version increased for intermediate reverberation times and decreased for high reverberation times, contrarily to the other datasets where the ELL varied monotonously with the reverberation level. In this experiment the room was virtual, created with the ray-tracing software CATT-Acoustic. The generation of the BRIRs is done with a random projection of rays across the room, a stochastic process. This could explain why the ELLs do not vary monotonously, as the detailed pattern of reflections randomly changes between BRIRs.

The Kokabi version was less accurate than the other versions for the datasets from Lavandier and Culling (2008) and the third experiment of Collin and Lavandier (2013), which involved high levels of reverberation. This is in accordance with the conclusions of Kokabi et al. (2018), who developed their linear regression based on long-term IACC at low and intermediate reverberation levels only. Our results indicate that the efficiency of this regression decreases at high levels of reverberation. It is interesting to note that this model version gave slightly better predictions than the others at very low reverberation levels, (in other words some conditions of the first and fourth experiments of Collin and

Lavandier (2013)). This could suggest that the linear regression might be appropriate for low reverberation levels. Alternatively it could indicate that using a shorter decay duration is more relevant then. Note however that low reverberation levels correspond to situations in which temporal smearing of the target will probably have negligible effect on intelligibility (Lavandier & Culling, 2008), greatly limiting the interest for the early-late separation to start with.

The model version based on short-term IACC gave similar performances to the room-independent version, except for the dataset from Lavandier and Culling (2008). The IACC version then estimated ELL shorter than the room-independent ELL, leading to a large overestimation of the effect of the temporal smearing of the target speech. Note that the method is binaural and could not be tested on the dataset from Cueille et al. (2022) also involving high levels of reverberation. Overall the estimated ELLs were always very short for this method (below 12 ms, Table 3). This would imply that only the direct sound is considered useful and most of the early speech reflections are considered detrimental, which is not appropriate as early speech reflections usually improve speech intelligibility.

Overall, the room-independent and Abel versions of the model gave the best predictions and were accurate for all datasets. The Abel version was more accurate than the room-independent version for one dataset with high level of reverberation, but less accurate for another dataset involving a virtual room. Even if Leclère et al. (2015) showed that optimizing the ELL as a function of the room/BRIR could improve intelligibility predictions, the room-independent version of the model can be preferred to the Abel version, because it is simpler to implement and has good performance across the wide range of conditions included in the seven datasets tested here. The Kokabi and IACC versions were not able to describe all datasets with sufficient accuracy due to underestimating, respectively overestimating, the effect of the temporal smearing of the target speech for some of the datasets with high levels of reverberation.

The internal noise only affected the predictions for the two datasets involving HI listeners (Cueille et al., 2022; Vicente et al., 2021). The version using the external sound broadband level was slightly more accurate for the dataset from Vicente et al. (2021), while the version using the external spectrum was more accurate with the dataset from Cueille et al. (2022). The internal noise is primarily affected by the audiograms of the listeners, and the main difference between the listeners of the two datasets is the degree of hearing loss of the HI listeners, the HI listeners of Cueille et al. (2022) having a higher degree of hearing loss. The internal noise taking the broadband level as input could be too high for the HI listeners with a high degree of hearing loss, while the internal noise using the spectrum would be too low for the HI listeners with a low degree of hearing loss. This assumption is supported by the individual predictions of the dataset from Cueille et al. (2022) (Supplemental Figure 5), which show



that considering the broadband level largely overestimates the SRTs of some of the HI listeners. It is also in agreement with the results of Vicente et al. (2020), who found that taking the broadband level as input allowed the model to give accurate predictions for HI listeners with a low degree of hearing loss. Another difference between the two datasets used for testing the model versions is the listening mode used for data collection: Monaural (Cueille et al., 2022) vs. binaural (Vicente et al., 2021). An interaction between the internal noise implementation and the listening mode cannot be ruled out to explain the differences observed. Overall, it appears at this stage that we cannot tell which internal noise calculation method is better. This requires further investigation, considering yet other datasets.

### *Comparison with Other Models from the Literature*

The main advantage of the proposed model is that it allows for predictions in a large variety of conditions, taking into account the effects of binaural hearing, hearing impairment, reverberation, and interferer modulations. There exist other binaural models which take into account the effect of hearing impairment by taking the listener's audiogram as input, such as the one from Beutelmann et al. (2010). This model was tested by Beutelmann et al. (2010) and Ewert et al. (2017) with HI and NH listeners, as well as modulated and stationary noise. However, it cannot take into account the effect of the temporal smearing of the target speech because it considers the whole target speech as useful for speech intelligibility.

There also exist other binaural models which allow to predict the effect of the temporal smearing of the target speech caused by reverberation, such as the ones developed by van Wijngaarden and Drullman (2008), Chabot-Leclerc et al. (2016), Andersen et al. (2018) and Rennies et al. (2011). However, the models of van Wijngaarden and Drullman (2008), Chabot-Leclerc et al. (2016) and Andersen et al. (2018) work only for NH listeners. Furthermore, Andersen et al. (2018) found that the model of Chabot-Leclerc et al. (2016) largely overestimates SRM. The models proposed by Rennies et al. (2011) and Rennies et al. (2019) are extensions of the model of Beutelmann et al. (2010) which can predict the temporal smearing of the target speech. One of those extensions does an early-late separation of the BRIR as in the proposed model. Like the original Beutelmann et al. (2010) model, these extensions should in principle work for HI listeners. But contrary to the model of Beutelmann et al. (2010), they cannot predict the effect of dip listening because they do not consider the modulations of the interfering masker. Furthermore, to the best of our knowledge, they were so far only tested for NH listeners and stationary noise.

When tested in different conditions than those considered here, the model from Beutelmann et al. (2010) led to similar individual prediction performances than the proposed model, with a correlation of 0.78, and a mean error of 3 dB. Among the three model versions from Rennies et al. (2011), the more

accurate gave a correlation of 0.97 and a mean error of 1 dB for averaged predictions. The model from Andersen et al. (2018) gave across-conditions correlations between 0.97 and 0.98 considering averaged data across listeners. It would be interesting to verify that the model from Rennies et al. (2011) work correctly with HI listeners in reverberant conditions. Unfortunately, the performance statistics of this model cannot be directly compared to the ones obtained with the proposed model, as they were obtained with datasets involving other speech materials and different conditions. It would be interesting to test the proposed model along with the models from the literature on the same datasets in order to compare their performances.

The models proposed during the Clarity Challenge allow to predict the performances of HI listeners with hearing aids in reverberant conditions (Barker et al., 2022). This is also the case for the model from Schädler et al. (2018), which follows a similar machine learning approach. Some of these models also have the advantage of being nonintrusive, necessitating only the combination of the speech and noise signals at the two ears, while the proposed model requires the separated target and noise signals as well as the target and masker BRIRs as input, limiting its range of application. However, some of the more accurate models of the Clarity Challenge as well as the model from Schädler et al. (2018) necessitate training with speech and listening conditions similar to the evaluation material beforehand. Furthermore, the Clarity Challenge made no clear separation between the training and test materials, which creates the risk that overfitted machine learning models were rewarded.

### *Limitations of the Proposed Model*

The dataset from Cueille et al. (2022) is slightly less accurately predicted than the other datasets tested here, with larger prediction errors. This might be partly explained by the external sound level used in the model. During the experiment, the target speech was set at a fixed sound level of 50 dB SPL for the NH listeners, while the interfering noise (which was louder than the speech in most conditions that led to negative SRTs) varied in level depending on the condition. In the model, the external sound level was fixed to the target speech level for all conditions. Even if the higher noise levels would have been a better approximation of the external sound levels at negative SNRs, these levels cannot be used as a model input because they are fixed by the SRT that corresponds to the model output. The use of the target level could have led to an underestimation of the internal noise of the HI listeners in some conditions and affected the predictions.

The experiment from Lavandier and Culling (2008) is also not perfectly predicted, with a correlation coefficient of 0.85 between measured and predicted SRTs. The room-independent version of the model underestimates speech intelligibility in the conditions with an anechoic target, and overestimates it in the conditions with a moderately

reverberant target. This is similar to the predictions obtained with *leclere2015*. According to the model predictions, this intelligibility increase (SRT decrease) between the anechoic and moderately reverberant conditions would be due to coloration which would have influenced better-ear listening (Leclère et al., 2015). However, the listeners did not seem to take any advantage of this coloration. This effect remains lower than 1 dB.

The model slightly overestimates the dip listening benefit for experiment 4 of Collin and Lavandier (2013) and the datasets of Cueille et al. (2022) and Vicente et al. (2021). This is something that was already observed by Vicente et al. (2020). It is likely a misprediction from the better-ear listening component, as the dip listening overestimation is present even in the conditions where there is no binaural unmasking. For example, in the dataset of Vicente et al. (2021), with a target azimuth of 20°, a dip listening benefit of 2.7 dB is observed in the data, while the model predicted a benefit of 4.1 dB, of which 4 dB are due to better-ear listening. Overall, this overestimation of the dip listening benefit remains smaller than 2 dB for all tested datasets. This error size is similar to the one obtained for a different dataset with the original approach of the short-term SII (Rhebergen & Versfeld, 2005). To further reduce this error, one could try to consider frequency-dependent durations for the temporal windows used for the time-frame decomposition of the masker signal, the fixed-duration windows used here being only an approximation of a more complex frequency-dependent decomposition of the signals proposed by Rhebergen and Versfeld (2005). Note that this might not be trivial to implement in our current modelling framework, in which the time-frame decomposition is introduced before the frequency band analysis. One could also refine the model to consider forward spread of masking, as it was shown to produce better prediction of dip listening for noise maskers with large silent gaps and abrupt offsets, even if it did not improve prediction for sinusoidally-modulated noises without silent periods in Rhebergen et al. (2006).

For experiment 4 of Collin and Lavandier (2013) the model overestimates the SRM in the symmetrical condition for the 2-voice modulated maskers. This was already observed with *vicente2020* and could be due to binaural sluggishness not being taken into account for the better-ear listening component of the model (Vicente & Lavandier, 2020).

A drawback of the proposed model is its intrusiveness. As explained above, models such as those of the Clarity Challenge only need the noisy speech signals at the two ears, while most other models need either only the separated speech and noise signals at the ears or only the BRIRs. The proposed model needs the raw, unconvolved target and masker signals, as well as the corresponding BRIRs. This required data explains why the model was yet only tested with datasets available in our laboratories.

The proposed model also cannot predict conditions where the late reflections of the target speech are useful to

intelligibility, unlike the model from Rennies et al. (2019). Interestingly, Rennies et al. (2022) found that HI listeners seemed generally less able to “ignore” the direct sound than NH listeners in these conditions, with an individual variability that was not completely explained by the listener audiogram. It could be interesting to further test the approach proposed by Rennies et al. (2019) with HI listeners. However such conditions are uncommon, and it is unclear whether this approach would benefit the proposed model.

As for *vicente2020*, the proposed model does not consider all aspects of hearing loss. Due to only taking the audiogram of the listeners as input, it does not include several factors which can affect speech intelligibility. For example, it has been shown HI listeners can have reduced compression (Oxenham & Bacon, 2003) and frequency selectivity (Strelcyk & Dau, 2009), both of which can impact temporal processing, including forward masking and dip listening abilities. HI listeners can also experience a broadening of their auditory filters (Glasberg & Moore, 1986) or a loss of ITD sensitivity (King et al., 2014), which could affect SRM. These factors can explain why the models do not fully predict the variability among HI listeners (Supplemental Figures 5 and 7) and instead focus on the differences between the NH and HI groups.

The model does not account for the effect of age on speech intelligibility (Füllgrabe et al., 2015). Age differences between listener groups may contribute to intelligibility score differences between NH and HI listeners (in particular, it could have been the case in the dataset from Vicente et al. (2021) where the listener groups were not matched in age).

Note also that the proposed model can only predict differences between conditions in a given dataset. As for *vicente2020* and *leclere2015* (Lavandier et al., 2022), nothing is implemented to account for the effect of a difference in speech material between two datasets.

Finally, the datasets considered here only allow to test a few combinations of individual effects (reverberation, hearing impairment, SRM, dip listening). It could be interesting to further test the model with a dataset involving all these effects simultaneously, but this would require a rather large number of tested conditions.

## Conclusion


A combination of two binaural speech intelligibility models *vicente2020* and *leclere2015* is proposed. The new model was as accurate as the two original models for the datasets that they could predict. It was also accurate for the datasets involving at the same time a temporally smeared target and a modulated noise for HI and NH listener groups, something that *vicente2020* and *leclere2015* cannot do. The four tested methods of BRIR separations had varying prediction accuracy. Two of them gave comparable predictions: The room-independent version, for which the ELL is fixed to 30 ms (Leclère et al., 2015), and the version for which the ELL is


defined as the moment when the diffuse reverberation tail begins (Abel & Huang, 2006). The two other versions based on the IACC (long-term or short-term) were less accurate for at least one of the tested datasets. The room-independent version might be preferable due to its simpler implementation. Overall, the two methods of internal noise calculation that were tested gave similarly accurate predictions. The method developed in *vicente2020* was slightly more accurate for HI listeners with a moderate hearing loss and binaural listening, while the new proposed method was more accurate for HI listeners with high hearing loss and monaural listening.

## Acknowledgments

The authors thank Christer P. Volk for providing the toolbox necessary to estimate the ELL based on the mixing time detection. The data from the experiments with human participants are taken from previously published studies.

## ORCID iDs

Raphael Cueille  <https://orcid.org/0000-0003-1308-0591>

Mathieu Lavandier  <https://orcid.org/0000-0002-2976-4608>

## Funding

The authors received no financial support for the research, authorship, and/or publication of this article: This work was conducted within the LabEx CeLyA (Grant No. ANR-10-LABX-0060). It has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101129903.

## Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Data Availability Statement

The model developed here is being made publicly available within the Auditory Modeling Toolbox (AMT 1.1; Majdak et al., 2022), along with the code, data and signals used to run the model for an example experiment, as done by Lavandier et al. (2022).

## Supplemental Materials

Supplemental material for this paper is available online. See supplemental materials at [URL will be inserted by SAGE] for the figures comparing the predictions of the proposed model to those of *vicente2020* and *leclerc2015*, as well as those presenting the individual predictions.

## References

- Abel, J. S., & Huang, P. (2006). A simple, robust measure of reverberation echo density. In *Audio engineering society convention 121*, Audio Engineering Society.
- Andersen, A. H., de Haan, J. M., Tan, Z. -H., & Jensen, J. (2016). Predicting the intelligibility of noisy and nonlinearly processed binaural speech. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(11), 1908–1920. <https://doi.org/10.1109/TASLP.2016.2588002>
- Andersen, A. H., de Haan, J. M., Tan, Z. -H., & Jensen, J. (2018). Refinement and validation of the binaural short time objective intelligibility measure for spatially diverse conditions. *Speech Communication*, 102, 1–13. <https://doi.org/10.1016/j.specom.2018.06.001>
- ANSI S3.5 (1997). "Methods for calculation of the speech intelligibility index" American National Standards Institute, New York.
- Barker, J., Akeroyd, M., Cox, T. J., Culling, J. F., Firth, J., Graetzer, S., Griffiths, H., Harris, L., Naylor, G., Podwinska, Z. et al (2022). The 1st clarity prediction challenge: A machine learning challenge for hearing aid intelligibility prediction. In *Interspeech 2022* (pp. 3508–3512), <https://doi.org/10.21437/Interspeech.2022-10821>.
- Bench, J., Kowal, Å., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, 13(3), 108–112. <https://doi.org/10.3109/03005367909078884>
- 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. <https://doi.org/10.1121/1.2202888>
- Beutelmann, R., Brand, T., & Kollmeier, B. (2010). Revision, extension, and evaluation of a binaural speech intelligibility model. *The Journal of the Acoustical Society of America*, 127(4), 2479–2497. <https://doi.org/10.1121/1.3295575>
- Bronkhorst, A., & 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*, 83(4), 1508–1516. <https://doi.org/10.1121/1.395906>
- Chabot-Leclerc, A., MacDonald, E. N., & Dau, T. (2016). 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.
- 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. <https://doi.org/10.1121/1.4812248>
- 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. <https://doi.org/10.1098/rsos.210342>
- Culling, J. F., Hawley, M. L., & Litovsky, R. Y. (2005). 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), 552–552. <https://doi.org/10.1121/1.5092204>
- Dillon, H. (2012). *Hearing Aids* (2nd ed). New York: Chap. 10 (Thieme).
- Durlach, N. I. (1963). Equalization and cancellation theory of binaural masking-level differences. *The Journal of the Acoustical Society of America*, 35(8), 1206–1218. <https://doi.org/10.1121/1.1918675>

- Ewert, S. D., Schubotz, W., Brand, T., & Kollmeier, B. (2017). Binaural masking release in symmetric listening conditions with spectro-temporally modulated maskers. *The Journal of the Acoustical Society of America*, 142(1), 12–28. <https://doi.org/10.1121/1.4990019>
- 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. <https://doi.org/10.1121/1.400247>
- Füllgrabe, C., & Moore, B. C. (2017). 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. <https://doi.org/10.3389/fnagi.2014.00347>
- Füllgrabe, C., Moore, B. C., & Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. *Frontiers in Aging Neuroscience*, 6, 347.
- Glasberg, B. R., & Moore, B. C. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *The Journal of the Acoustical Society of America*, 79(4), 1020–1033.
- 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. <https://doi.org/10.1177/1084713811424885>
- Helfer, K. S., & Wilber, L. A. (1990). Hearing loss, aging, and speech perception in reverberation and noise. *Journal of Speech, Language, and Hearing Research*, 33(1), 149–155. <https://doi.org/10.1044/jshr.3301.149>
- Houtgast, T., & Steeneken, H. J. (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. <https://doi.org/10.1121/1.392224>
- IEEE (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics* 17(3), 225–246. <https://doi.org/10.1109/TAU.1969.1162058>
- Jansen, S., Luts, H., Wagener, K. C., Kollmeier, B., Del Rio, M., Dauman, R., James, C., Fraysse, B., Vormès, E., Frachet, B. et al (2012). Comparison of three types of French speech-in-noise tests: A multi-center study. *International Journal of Audiology*, 51(3), 164–173. <https://doi.org/10.3109/14992027.2011.633568>
- King, A., Hopkins, K., & Plack, C. J. (2014). The effects of age and hearing loss on interaural phase difference discrimination. *The Journal of the Acoustical Society of America*, 135(1), 342–351.
- 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. <https://doi.org/10.1121/1.5078598>
- Lavandier, M., & Best, V. (2020). Modeling binaural speech understanding in complex situations. In J. Blauert & J. Braasch (Eds.), *The technology of binaural understanding* (pp. 547–578). (Springer Nature Switzerland), [https://doi.org/10.1007/978-3-030-00386-9\\_19](https://doi.org/10.1007/978-3-030-00386-9_19).
- 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. <https://doi.org/10.1121/1.2871943>
- 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. <https://doi.org/10.1121/1.3268612>
- 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. <https://doi.org/10.1121/1.3662075>
- Lavandier, M., Mason, C. R., Baltzell, L. S., & Best, V. (2021). Individual differences in speech intelligibility at a cocktail party: A modeling perspective. *The Journal of the Acoustical Society of America*, 150(2), 1076–1087. <https://doi.org/10.1121/1.0005851>
- Lavandier, M., Vicente, T., & Prud'homme, L. (2022). A series of SNR-based speech intelligibility models in the auditory modeling toolbox. *Acta Acustica*, 6, 20.
- 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. <https://doi.org/10.1121/1.4921028>
- 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.
- Lochner, J., & Burger, J. (1964). The influence of reflections on auditorium acoustics. *Journal of Sound and Vibration*, 1(4), 426–454. [https://doi.org/10.1016/0022-460X\(64\)90057-4](https://doi.org/10.1016/0022-460X(64)90057-4)
- Majdak, P., Hollomey, C., & Baumgartner, R. (2022). AMT 1. x: A toolbox for reproducible research in auditory modeling. *Acta Acustica*, 6, 19. <https://doi.org/10.1051/aacus/2022011>
- Miller, G. A., & Licklider, J. C. (1950). The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America*, 22(2), 167–173. <https://doi.org/10.1121/1.1906584>
- Moore, B. C., & Glasberg, B. R. (2004). A revised model of loudness perception applied to cochlear hearing loss. *Hearing research*, 188(1-2), 70–88.
- Oxenham, A. J., & Bacon, S. P. (2003). Cochlear compression: Perceptual measures and implications for normal and impaired hearing. *Ear and Hearing*, 24(5), 352–366.
- 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.
- Raake, A., & Katz, B. F. (2006). SUS-based method for speech reception threshold measurement in French. In *LREC* (pp. 2028–2033).
- 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. <https://doi.org/10.1121/1.3641368>
- Rennies, J., Warzybok, A., Brand, T., & Kollmeier, B. (2019). Measurement and prediction of binaural-temporal integration of

- speech reflections. *Trends in Hearing*, 23, 1–22. <https://doi.org/10.1177/2331216519854267>
- Rennies, J., Warzybok, A., Kollmeier, B., & Brand, T. (2022). Spatio-temporal integration of speech reflections in hearing-impaired listeners. *Trends in Hearing*, 26, 23312165221143901.
- Rhebergen, K. S., & Versfeld, N. J. (2005). 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.
- Rhebergen, K. S., Versfeld, N. J., Dreschler, W. et al (2006). 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.
- Schädler, M. R., Warzybok, A., & Kollmeier, B. (2018). Objective prediction of hearing aid benefit across listener groups using machine learning: Speech recognition performance with binaural noise-reduction algorithms. *Trends in Hearing*, 22, 2331216518768954.
- Strelcyk, O., & Dau, T. (2009). 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.
- Taal, C. H., Hendriks, R. C., Heusdens, R., & Jensen, J. (2011). An algorithm for intelligibility prediction of time–frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7), 2125–2136. <https://doi.org/10.1109/TASL.2011.2114881>
- van Wijngaarden, S. J., & Drullman, R. (2008). Binaural intelligibility prediction based on the speech transmission index. *The Journal of the Acoustical Society of America*, 123(6), 4514–4523. <https://doi.org/10.1121/1.2905245>
- Vicente, T., Buchholz, J. M., & Lavandier, M. (2021). Modelling binaural unmasking and the intelligibility of speech in noise and reverberation for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 150(5), 3275–3287. <https://doi.org/10.1121/10.0006736>
- Vicente, T., & Lavandier, M. (2020). Further validation of a binaural model predicting speech intelligibility against envelope-modulated noises. *Hearing Research*, 390, 107937. <https://doi.org/10.1016/j.heares.2020.107937>
- 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. <https://doi.org/10.1121/10.0002660>
- 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. <https://doi.org/10.1121/1.5026788>