A binaural model implementing an internal noise to predict the effect of hearing impairment on speech intelligibility in non-stationary noises

Thibault Vicente, Mathieu Lavandier, and Jörg M. Buchholz

Citation: The Journal of the Acoustical Society of America 148, 3305 (2020); doi: 10.1121/10.0002660
View online: https://doi.org/10.1121/10.0002660
View Table of Contents: https://asa.scitation.org/toc/jas/148/5
Published by the Acoustical Society of America

ARTICLES YOU MAY BE INTERESTED IN

A harmonic-cancellation-based model to predict speech intelligibility against a harmonic masker
The Journal of the Acoustical Society of America 148, 3246 (2020); https://doi.org/10.1121/10.0002492

Effect of masks on speech intelligibility in auralized classrooms
The Journal of the Acoustical Society of America 148, 2878 (2020); https://doi.org/10.1121/10.0002450

Relationship between sensitivity to temporal fine structure and spoken language abilities in children with mild-to-moderate sensorineural hearing loss
The Journal of the Acoustical Society of America 148, 3334 (2020); https://doi.org/10.1121/10.0002669

Pitch perception at very high frequencies: On psychometric functions and integration of frequency information
The Journal of the Acoustical Society of America 148, 3322 (2020); https://doi.org/10.1121/10.0002668

Assessing the benefit of acoustic beamforming for listeners with aphasia using modified psychoacoustic methods
The Journal of the Acoustical Society of America 148, 2894 (2020); https://doi.org/10.1121/10.0002454

Measurements of ear-canal cross-sectional areas from live human ears with implications for wideband acoustic immittance measurements
The Journal of the Acoustical Society of America 148, 3042 (2020); https://doi.org/10.1121/10.0002358
A binaural model implementing an internal noise to predict the effect of hearing impairment on speech intelligibility in non-stationary noises

Thibaut Vicente, Mathieu Lavandier, and Jörg M. Buchholz

ABSTRACT:
A binaural model predicting speech intelligibility in envelope-modulated noise for normal-hearing (NH) and hearing-impaired listeners is proposed. The study shows the importance of considering an internal noise with two components relying on the individual audiogram and the level of the external stimuli. The model was optimized and verified using speech reception thresholds previously measured in three experiments involving NH and hearing-impaired listeners and sharing common methods. The anechoic target, in front of the listener, was presented simultaneously through headphones with two anechoic noise-vocoded speech maskers (VSs) either co-located with the target or spatially separated using an infinite broadband interaural level difference without crosstalk between ears. In experiment 1, two stationary noise maskers were also tested. In experiment 2, the VSs were presented at different sensation levels to vary audibility. In experiment 3, the effects of realistic interaural time and level differences were also tested. The model was applied to two datasets involving NH listeners to verify its backward compatibility. It was optimized to predict the data, leading to a correlation and mean absolute error between data and predictions above 0.93 and below 1.1 dB, respectively. The different internal noise approaches proposed in the literature to describe hearing impairment are discussed.

I. INTRODUCTION
Having two ears can be very helpful to improve speech intelligibility (SI) in a noisy environment. Spatial release from masking (SRM) is the ability to benefit from a spatial separation between target and masker sources using binaural cues, the interaural level difference (ILD) and interaural time difference (ITD), to improve SI [e.g., Bronkhorst and Plomp (1988) and Culling et al. (2004)]. When a target speaker is spatially separated from the competing masking noise(s), differences in ILD and ITD are observed between the signals produced by the target and masking sources at the ears of the listener. The differences in ILD result in different signal-to-noise ratios (SNRs) at the two ears. Better-ear listening reflects our ability to exploit the higher SNR across ears to better understand the target. Binaural unmasking is a mechanism that utilizes the differences in ITD to improve SI. Durlach (1963) modelled the latter by assuming that a listener is able to internally equalize the masking signals at the ears—by applying a gain and a delay at one ear to compensate for the masker ILD/ITD—and partially cancel this masker to improve the internal SNR and, consequently, SI. This theory is called Equalization-Cancellation (EC) theory.

The SRM is typically reduced for hearing-impaired (HI) listeners because the benefit of these auditory mechanisms is substantially degraded by hearing loss (HL), which leads to more difficulties with understanding speech in noisy environments [e.g., see Glyde et al. (2011) for a review]. Hearing impairment can be characterized by an elevated hearing threshold, as measured by a pure-tone audiogram, which results in reduced audibility of important speech and masker information, and thus degrades SI (Rana and Buchholz, 2018b). Reduced audibility has a direct effect on the benefit provided by better-ear listening, since the speech signal at the ear with the better SNR may not be audible or only partially audible (Glyde et al., 2013). In addition, other aspects related to HL can impair SI in noise, such as reduced sensitivity to the temporal fine structure of a signal [e.g., Moore (2008)]. The latter has direct implications for the ITD sensitivity in HI listeners (Füllgrabe and Moore, 2017). Whereas normal-hearing (NH) listeners are very sensitive to
changes in ITD up to a frequency of around 1.3 kHz, HI listeners show reduced sensitivity to ITD with a decrease in the upper cut-off frequency (King et al., 2014; Neher et al., 2011). A reduced ITD sensitivity has been linked to a reduced benefit provided by binaural unmasking (Santurette and Dau, 2012).

In order to predict SI in noise, several models have been proposed that simulate the relevant auditory signal processing with more or less accuracy. A simplistic approach is to assume that the auditory system is linear and filters the signals per frequency band. In this regard, the monaural and binaural models developed by Taal et al. (2011) and Andersen et al. (2016) predict SI for NH listeners using a third-octave-band decomposition. Other monaural (Jørgensen et al., 2013; Rhebergen and Versfeld, 2005) and binaural (Chabot-Leclerc et al., 2016; Collin and Lavandier, 2013; Jelfs et al., 2011; Wan et al., 2014) models for NH listeners have been developed using a gammatone filterbank, which consists of linear band-pass filters that mimic the shape of the auditory filters. The SI index (SII) (ANSI, 1997) is a monaural model that can be applied using four different frequency band distributions to predict SI for NH and HI listeners. A number of SI models implemented a non-linear behavior of the auditory periphery. For instance, Relano-Iborra et al. (2019) developed a model for monaural listening that is yet to be validated for HI listeners, and Bruce et al. (2013) and Scheidiger et al. (2018) proposed monaural models that take into account the different effects of inner hair cell (IHC) and outer hair cell (OHC) losses.

A number of binaural models have been proposed to predict the effect of hearing impairment on SI using a gammatone filterbank (Beutelmann and Brand, 2006; Beutelmann et al., 2010; Lavandier et al., 2018). They use the listener’s audiograms to implement an internal noise that is then considered as an additional masking noise (like in the [SII]). The model from Beutelmann et al. (2010) was able to explain 78% of the variation of the data found across a range of (reverberant) conditions, but the predicted SRTs were on average 3.4 dB lower than the data, and predictions were generally less accurate for HI listeners. The model by Lavandier et al. (2018) was able to accurately predict NH and HI data measured in two (anechoic) spatial configurations, but applied a model parameter (associated with the broadband level of the internal noise) that had to be chosen differently for the NH and HI listeners to obtain accurate predictions; effectively resulting in two different model versions. Furthermore, their approach divided the listeners only in two groups (NH and HI), even though there are many different degrees (and types) of HL, which would again require separate model versions.

In the present study, the model developed by Lavandier et al. (2018) is revised, so that a single model can be used to predict SI for listeners with various degrees of HL instead of separate models. It was accomplished by modifying the implementation of the internal noise, which consists now of two components: one related to elevated hearing thresholds and the other related to the effect of external stimulus level on SI. This approach is in line with the findings from Bernstein and Trahiotos (2008), who provide an overview of the literature on the concept of internal noise before conducting experiments to measure detection thresholds for a tone in noise to characterize the internal noise in NH listeners. In line with their literature review, their results suggest that the internal noise would consist of two components. The first component is stimulus independent and determines the absolute threshold (i.e., serving as ‘noise floor’). The second component is stimulus-dependent, with its level increasing when the external noise level increases following a dB-for-dB rule.

The revised model is optimized and verified here using data from three experiments involving NH and HI listeners (Rana and Buchholz, 2016, 2018a,b), as well as data measured with only NH listeners (Collin and Lavandier, 2013; Lavandier et al., 2012) to verify its backward compatibility. Besides addressing the model performance, the discussion provides an in-depth analysis of the internal noise implementation and a comparison with other models proposed in the literature that consider an internal noise.

II. MODEL DESCRIPTION
A. Original model developed for NH listeners

A block diagram of the proposed SI model is shown in Fig. 1, with the components of the original model (Collin and Lavandier, 2013; Vicente and Lavandier, 2020) highlighted with the black font. The target and combined masker signals at the listener’s ears, equalized to the same broadband level (mean across ears), are taken as inputs to the model. The target characteristics (i.e., magnitude spectra at each ear and ITDs) are averaged across time, computed only once on the whole signal to get their long-term values. This avoids that a short pause between words leads to a very low SNR and thus a poor predicted SI, even though it provides relevant information for the listener. The masker characteristics are computed as a function of time for the model to predict the ability of a listener to understand speech in the dips of the masker’s envelope (“dip listening,” Festen and Plomp (1990)).

Based on the incoming signals, the SNR at the better ear and the binaural unmasking advantage are computed per time frame and frequency band combining the target’s long-term characteristics with the masker’s short-term characteristics (i.e., magnitude spectra at each ear, ITDs, and interaural coherence). To compute the SNR at the two ears, the masker signals are segmented using 24-ms half-overlapping Hann windows, so that “monaural” dip listening can be predicted. To compute the binaural unmasking advantage, 300-ms half-overlapping Hann windows are used so that binaural sluggishness can be taken into account (Hauth and Brand, 2018; Vicente and Lavandier, 2020). The frequency analysis is realized by applying two Gammatone filters per equivalent rectangular bandwidth, covering a
frequency range from 30 Hz to the closest gammatone filter center frequency below the half of the sampling frequency, i.e., 9.6, 19.9, and 25.1 kHz for the signals sampled at 20 kHz (Lavandier et al., 2012), 44.1 kHz (Rana and Buchholz, 2016, 2018a,b), and 48 kHz (Collin and Lavandier, 2013), respectively.

The better-ear SNR is computed choosing the higher SNR between the left and right ears, with a ceiling SNR at 20 dB to avoid that the SNR goes to infinity in the masker’s dips. The binaural unmasking advantage is computed by applying the binaural masking level difference formula from Culling et al. (2005). Their values are averaged across time, integrated across frequency using a SII weighting [derived from the band importance function in Table I of ANSI (1997)], and added to obtain a binaural ratio. Differences between binaural ratios can be directly compared to differences between speech reception thresholds (SRTs, SNR at 50% of the target speech is intelligible) measured in listening tests.

In order to derive predicted SRTs, the binaural ratios are first inverted, so that a higher binaural ratio reflects better SI. Then, for each experiment, they are offset to fit the data by subtracting their mean and adding the average measured SRT across conditions and listeners, which 50% of the target speech is intelligible) measured in listening tests.

In order to derive predicted SRTs, the binaural ratios are first inverted, so that a higher binaural ratio reflects better SI. Then, for each experiment, they are offset to fit the data by subtracting their mean and adding the average measured SRT across conditions and listeners, which 50% of the target speech is intelligible) measured in listening tests.

B. Extension of the model to HI listeners

In order to quantify the effect of reduced audibility, Lavandier et al. (2018) applied a number of changes to the original model of Collin and Lavandier (2013), using the implementation suggested by Cubick et al. (2018). These changes are highlighted with the grey font in Fig. 1. First, the model input signals are calibrated to the sound level [in dB of sound pressure level (SPL)] used during the experiment (i.e., the sound level of the masker here, that was fixed during the adaptive measurements). This means that any amplification applied to the stimuli in order to compensate for the listener’s HL is considered by the model. To take into account the HL, an internal noise is implemented at each ear with a spectrum that matches the individual audiograms. To compute the SNR at the better ear, the SNR at each ear is determined by the higher level between the external and internal noise and limited to 20 dB. The highest SNR across ears is selected as the better-ear SNR. The binaural unmasking advantage is computed only if the masker and target levels are above the internal noise levels at both ears.

The differences applied in the present study to the model of Lavandier et al. (2018) are solely related to the implementation of the internal noise. In this regard, three assumptions are made. (1) The overall level of the external stimuli is approximated by the known masker level, assuming that the broadband SNR is below 0 dB. (2) The HL of the listener can be split into proportions $\eta$ and $1 - \eta$ to reflect the different contributions of OHC and IHC loss, with $\eta$ being identical at all frequencies and for all listeners. (3) The maximum value allowed for the estimated OHC loss is 57.6 dB (Moore and Glasberg, 2004) and the loss above this value contributes solely to the estimated IHC loss. The estimated OHC and IHC losses are interpolated at the model’s center frequencies between the lower and upper center frequency of the audiograms (250 Hz and 8000 Hz, respectively) and extrapolated otherwise, using a logarithmic frequency scale. The level (in dB SPL) of the internal noise $N_{\text{int}}$ in each frequency band is then calculated using the following formula:

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

where $n$ is the center frequency of the $n$th frequency band and $N$ is the long-term broadband level of the external masker averaged across ears. Given that the model takes into account the listener-individual amplification applied in the experiment, the term $N$ can vary across listeners (if they have different HL profiles leading to individual amplification). The term $T(n)$ refers to the standard pure-tone audiogram in dB HL, thus, $\eta T(n)$ and $(1 - \eta)T(n)$ are the
estimated contributions in dB HL related to OHC and IHC loss, respectively. The function $\Gamma(n)$ is the transformation to convert the HL from dB HL to dB SPL. This transformation results from the sum of the reference equivalent sound pressure levels for the THD 39 headphones used when measuring the audiograms (ISO 389-2, 1994) and nominal values for the transformation from 6 cc coupler to ear drum levels (Bentler and Pavlovic, 1989). The transformation follows the same interpolation/extrapolation as the audiograms (interpolated between 200 Hz and 6300 Hz, the range where values are available) to derive its values at the center frequencies of the model. The frequency-independent free parameters $B$, $N_{\text{lim}}$, and $\eta$ [highlighted in bold in Eq. (1)] were systematically varied in the present study for the experiments involving HI listeners, before being set at $-10$ dB, $83$ dB, and $0.7$, respectively (see Sec. II C).

The proposed implementation of the internal noise models a behaviour that resembles the different effects of OHC and IHC loss on SI. It is thereby assumed that the OHCs are mainly related to the audibility of the incoming sounds, while the IHCs are mainly related to their coding (e.g., Moore and Glasberg, 2004). Equation (1) at low sound levels $N$ [i.e., $N - N_{\text{lim}} + (1 - \eta)T(n) \ll B$] can be simplified to

$$N_{\text{int}}(n, N) \sim \Gamma(n) + \eta T(n) + 10 \log_{10}(10^{B/10} + 10^{(N - N_{\text{lim}} + (1 - \eta)T(n))/10})$$

$$\sim \Gamma(n) + \eta T(n) + B. \quad (2)$$

Hence, for soft sounds and poor audibility, only the part of the audiograms related to the OHCs [$\eta T(n)$] is considered. For high sound levels $N$ [i.e., $N - N_{\text{lim}} + (1 - \eta)T(n) \gg B$], Eq. (1) can be simplified to

$$N_{\text{int}}(n, N) \sim \Gamma(n) + \eta T(n) + 10 \log_{10}(10^{B/10} + 10^{(N - N_{\text{lim}} + (1 - \eta)T(n))/10})$$

$$\sim \Gamma(n) + \eta T(n) + (1 - \eta)T(n) + N - N_{\text{lim}}$$

$$\sim \Gamma(n) + T(n) + N - N_{\text{lim}}. \quad (3)$$

This indicates that, at high sound levels, in addition to the audibility issue related to the OHC loss, SI is further impaired by the IHC loss.

C. Model evaluation

The proposed model was evaluated on three datasets involving NH and HI listeners (Rana and Buchholz, 2016, 2018a,b), which considered the effects of HL, sensation level, spatial configuration of target and maskers, and masker temporal envelopes. Three indices of model performance were considered: the Pearson’s correlation coefficient $r$ between data and prediction, the mean absolute error (MeanErr) computed as the average across conditions of the absolute difference between data and prediction, and the maximum absolute error (MaxErr).

To quantify the improvement in prediction performance obtained with the current internal noise implementation compared to the implementation proposed by Lavandier et al. (2018), their models were applied here as reference ("Lav18 models"). The parameter value used to set the level of the internal noise for the NH listeners was $-11$ and $-22$ dB for the HI listeners (Lavandier et al., 2018). Compared to the original paper, instead of using a separate reference to convert the binaural ratios in predicted SRTs for each group of listeners, a common reference is used here for both groups, in order to evaluate whether the difference across groups can be predicted.

The backward compatibility with previous model versions was verified using two datasets involving only NH listeners. To compare predictions for a stationary noise masker, the model version and data from Lavandier et al. (2012) were considered ("Lav12 model"). To evaluate the effects associated with masker envelope modulations, the model version and data from Collin and Lavandier (2013) were considered ("Col13 model"). For a fair comparison, the components that are similar across models were implemented in the same way.

The Lav12 model follows the implementation presented in Sec. II A, except that no time frame analysis is applied, i.e., the better-ear SNR and the binaural unmasking advantage are computed on the long-term signal characteristics. The Col13 model follows the implementation presented in Sec. II A, but using the same 24-ms frame to compute the better-ear SNR and the binaural unmasking advantage [instead of using 24-ms and 300-ms frames proposed later by Vicente and Lavandier (2020)].

For all model predictions, the target signal was created by averaging between 60 and 128 sentence waveforms. All sentences were truncated to the shortest sentence duration before averaging. The duration of each masker signal was at least 2 min. All signals from the experiment of Rana and Buchholz (2016, 2018a,b) were both convolved with the impulse response of the (equalized) headphones used for data collection and measured on a 4128 C Brüel&Kjaer head and torso simulator. Target and masker signals were calibrated to the fixed sound level averaged across ears used for the masker in the experiments.

In order to find the best combination of parameters in Eq. (1), the free parameters $B$, $N_{\text{lim}}$, and $\eta$ were varied within the ranges $[-16; -8]$ dB, $[65; 85]$ dB, and $[0.6; 0.9]$ (Pieper et al., 2018), with the aim of simultaneously minimizing MeanErr and MaxErr and maximizing $r$. This optimization of the model performance was done only using the three datasets involving both NH and HI listeners (Rana and Buchholz, 2016, 2018a,b), and the best predictions were obtained for $B = -10$ dB, $N_{\text{lim}} = 83$ dB, and $\eta = 0.7$. The Spearman’s rank correlation coefficient $\rho$ was also computed between data and predictions for each experiment but not used as a criterion in the optimization stage.

III. RESULTS

The results shown below compare the SI data taken from the literature with the corresponding model predictions.
was not the case in Rana and Buchholz (2018a), where the audiograms were not available. This was simulated by convolving the anechoic stimuli with a head-related transfer function for frontal incidence position. The noises were equalized binaurally using simulated a HL at 0 dB HL at all frequencies. The target speech was a BKB-like corpus (Bench et al., 1979), consisting of 80 lists of 16 meaningful sentences containing between 4 and 7 words. It was always presented from the front of the listener simultaneously with two stationary speech-shaped noises (SSNs, envelope-unmodulated noises with the same spectrum as the target speech) or two noise-vocoded speech maskers (VSs, envelope modulated SSNs) either co-located with the target or spatially separated at ±90°. The frontal position was simulated by convolving the anechoic stimuli with a head-related transfer function for frontal incidence averaged across ears, resulting in diotic listening. The noises were presented at different sensation levels and the relative target level was adapted to derive the SRTs.

The predictions for the NH listeners of Rana and Buchholz (2016) and Rana and Buchholz (2018b) were computed simulating a HL at 0 dB HL at all frequencies for both ears because the audiograms were not available. This was not the case in Rana and Buchholz (2018a), where the individual audiograms for NH and HI subjects were available at each ear and used as model inputs. The predictions of the Lav18 models were not shown here because the figures would have been overloaded. However, the predictions are available online7 and the statistic performances and the main limitations (when applicable) are reported below for each experiment.

1. Experiment 1 of Rana and Buchholz (2016)

Ten young NH listeners aged between 23 and 42 years (mean age of 31.1 years) and 10 older HI listeners aged between 49 and 77 years (mean age of 66.9 years) were involved in this experiment. The four-frequency (0.5, 1, 2, 4 kHz) normal hearing (HL < 15 dB HL up to 6 kHz) or sensorineural HI with less than 10-dB-HL difference across ears at any audiometric frequency up to 4 kHz (symmetric HL). Moreover, the stimuli were anechoic and binaurally using equalized headphones. The target speech was played at four different sensation levels (0, 10, 20, and 30 dB SPL). Individual, non-ear-specific linear amplification was applied to the stimuli played to the HI listeners, following the National Acoustics Laboratories–Revised Profound (NAL-RP) prescription formula [Dillon (2012), Chap. 10], to compensate partly for their HL. The spatially separated configuration was designed by playing the left noise only through the left channel of the headphones and the right noise only through the right channel of the headphones, thus providing infinite broadband ILD.

Figure 2 shows the mean measured SRTs as a function of the masker type (circles), whereby the black and grey colours refer to the data collected with NH listeners and HI listeners, respectively. The open symbols represent the measured SRTs in the co-located configuration (“co-loc”) and the filled symbols show the SRTs for the spatially separated configuration (“separ”). The downward triangles correspond to the proposed model predictions and follow the same pattern of color and filling as the data. The good correlation with the data and low prediction errors (r = 0.98; ρ = 0.90; MeanErr = 1 dB; MaxErr = 2.1 dB) demonstrate that the model accurately predicts the effects across conditions and listener groups. In contrast, the Lav18 models are not able to predict well the difference between HI and NH listeners. The difference in measured SRTs across conditions between groups is about 7.2 dB, while the Lav18 models predict only 3.3 dB (and the proposed model predicts 6.4 dB). This discrepancy led also to worse model performances (r = 0.89; ρ = 0.76; MeanErr = 2.1 dB; MaxErr = 4.0 dB).

2. Experiment of Rana and Buchholz (2018a)

Ten young NH listeners aged between 20 and 30 years (mean age of 23.2 years) participated in this experiment along with 10 older HI listeners aged between 57 and 78 years (mean age of 70.3 years; 4FAHL = 29.1 ± 8.0 dB HL). Only the non-stationary VS maskers were considered. The spatially separated configuration was simulated as in Rana and Buchholz (2016), involving infinite broadband ILD. Target and maskers were filtered to individually equalize audibility across frequency for each listener and then played at four different sensation levels (0, 10, 20, and 30 dB SL) relative to their individual SRT in quiet. The filters were designed using detection thresholds in quiet obtained with a SSN filtered into nine frequency regions. The measurement was done separately for each listener and...
frequency region. The gain corresponding to the detection thresholds determined the filters that were then applied to the stimuli played to the listeners. Only the broadband condition was considered here (lowband, midband, and highband conditions were also tested in the original study). Amongst the 10 HI listeners, 1, 6, and 9 of them could not be tested at 10, 20, and 30 dB SL, respectively, due to loudness discomfort.

Figure 3 presents the measured SRTs as a function of masker sensation level in two different panels, one for each group of listeners. The open circles correspond to the SRTs measured for the condition where the target and masker were co-located and the filled circles represent the SRTs obtained for the spatially separated configuration. The dashed and solid lines correspond to the model predictions for the co-located and spatially separated conditions, respectively. The proposed model predicts the data with a similar accuracy as for the experiment of Rana and Buchholz (2018) (r = 0.98; ρ = 0.98; MeanErr = 1 dB; MaxErr = 3.1 dB). The predictions provided by the Lav18 models are as good as the proposed model (r = 0.97; ρ = 0.97; MeanErr = 1.4 dB; MaxErr = 3.3 dB), even though a general underestimation of the SRTs for the HI listeners can be observed, as well as an overestimation of 2.5 dB for the NH listeners at 0 dB SL.

3. Experiment of Rana and Buchholz (2018b)

Ten young NH participants aged between 25 and 41 years (mean age of 33.5 years) and 13 older HI listeners aged between 69 to 79 (mean age of 74 years; 4FAHL = 31 ± 8 dB HL) were involved in this experiment. Only the non-stationary VS maskers were tested and played at 60 dB SPL. Different amplification strategies were applied to the stimuli for the HI listeners to compensate partly for their HL, but only the individual, non-ear-specific linear (NAL-RP) amplification was considered here. Three different spatially separated configurations were tested in addition to a co-located configuration. The first one was spatialized using natural ILD and ITD, the second one involved natural ILD but no ITD, and the last one applied the same process as the two previous experiments with infinite broadband ILD. The three spatially separated configurations are in the following referred to as “Natural,” “ILD,” “Infinite ILD.”

Figure 4 presents the measured SRTs (filled circles) as a function of spatialization method. The black and grey symbols correspond to the mean measured SRTs of the NH listeners and HI listeners, respectively. The downward triangles show the proposed model predictions; using the same color pattern as the data. The same model accuracy as for the two previous experiments (Rana and Buchholz, 2016, 2018a) is obtained (r = 0.97; ρ = 0.95; MeanErr = 1 dB; MaxErr = 2.4 dB). Concerning the Lav18 models, the difference in predicted SRTs across spatial conditions between the HI and NH listeners is about 0.3 dB, which is about 5 dB less than what is observed in the data. This discrepancy led to a lower correlation and higher errors (r = 0.71; ρ = 0.79; MeanErr = 2.7 dB; MaxErr = 3.6 dB) than those obtained with the proposed model.

4. Individual differences between listeners

In order to investigate the accuracy of the model in predicting individual differences, a correlation analysis was conducted to compare the predicted SRTs with the measured SRTs for the HI listeners. For reference purposes, the correlation between the 4FAHL and the measured SRTs was also calculated. Only, the experiment involving the most HI listeners is considered here as an example, i.e., the one from Rana and Buchholz (2018b). The four correlation coefficients between the 4FAHL and measured SRTs (r = 0.92, 0.88, 0.82, 0.82) within each spatial condition (Co-located, Infinite ILD, Natural, and ILD) are higher than...
the coefficients between the predicted and measured SRTs \((r = 0.84, 0.84, 0.79, 0.78)\). However, the correlation between all predicted and measured SRTs across spatial conditions \((r = 0.88)\) is higher than the correlation between 4FAHL and measured SRTs \((r = 0.65)\). Hence, the 4FAHL explains slightly better the measured SRTs for a given spatial condition, but only the model is additionally able to describe the difference in SRTs across spatial conditions, which is not surprising as the 4FAHL was not designed to predict SRM and is independent of spatial condition.

### B. Dataset involving only NH listeners

In order to verify the backward compatibility of the proposed model, predictions for two experiments with only NH listeners are compared to previous model versions (Lav12 and Col13 models; see Sec. II C). As done above (and for the same reason), a NH listener was simulated here with 0 dB HL at all frequencies for both ears.

1. **Experiment 1 of Lavandier et al. (2012)**

   A meeting room was simulated through headphones, where a target was presented at 0.65 m and 25° of azimuth from the listener simultaneously with a SSN. The SSN was placed at one of two tested distances (0.65 and 5 m, referred to as “Near” and “Far”) and one of the three tested azimuths (−25°, 0°, 25° referred to as “Left,” “Front,” “Right”). Anechoic target and masker signals were convolved with binaural room impulse responses (BRIRs) recorded at each tested position. Spectral-envelope impulse responses (SEIRs) were also tested. They were short binaural impulse responses artificially obtained by removing the reverberation tail and ITD of the BRIRs, while preserving their spectral envelope at each ear (and the resulting ILD).

   Figure 5 presents the measured SRTs as a function of masker position. The black circles and grey squares show the SRTs measured with the BRIRs and SEIRs, respectively. The solid lines correspond to the predictions of the proposed model, while dashed lines present the predictions of the Lav12 model. The predictions of both models are very similar, the observed difference in each condition never exceeds 0.5 dB. The proposed model performance statistics \((r = 0.97; \rho = 0.95; \text{MeanErr} = 0.4 \text{ dB}; \text{MaxErr} = 0.6 \text{ dB})\) are therefore also very similar to the performances of the Lav12 model \((r = 0.98; \rho = 0.93; \text{MeanErr} = 0.3 \text{ dB}; \text{MaxErr} = 0.6 \text{ dB})\).

2. **Experiment 2 of Collin and Lavandier (2013)**

   Target and noise signals were presented diotically. The noise was a SSN either unmodulated or modulated with the broadband temporal envelope of 1, 2, or 4 simultaneous voices [the modulated noise changed from one target sentence to another while adaptively measuring the SRT, “variable” conditions, Collin and Lavandier (2013)].

   Figure 6 presents the measured SRTs as a function of masker modulation plotted as circles. The upward triangles display the predictions for the proposed model and the downward triangles correspond to the Col13 model. The predictions of the two models are almost identical with a maximal difference of 0.2 dB. The models also predict the data very well with identical performance statistics \((r = 0.93; \rho = 1; \text{MeanErr} = 0.5 \text{ dB}; \text{MaxErr} = 0.8 \text{ dB})\).

### IV. DISCUSSION

A. **Improvements obtained with the new model**

The proposed model has been optimized on three datasets involving NH and HI listeners. It accurately describes the effects of HL and presentation level on SI for NH and HI listeners using a single model version. This is a clear improvement over previous models.
improvement compared to the preceding models of Lavandier et al. (2018), which required different parameter values to define the internal noise for NH and HI listeners, thus effectively resulting in two different model versions. The improvement is achieved here by the new implementation of the internal noise that now depends on the external sound level and divides the listener’s HL into a proportion $\eta$ and $1 - \eta$ that roughly reflects the different effects of IHC and OHC loss on SI. The proposed model also predicts the two NH datasets as accurately as the previous models it is based on, which validates its backward compatibility. This highlights that the implementation of the internal noise considerably extends the scope of application of the model by taking HL into account, without compromising its previously demonstrated accuracy. Note that the model still needs to be tested on data not used to define its parameters, so that its predictive power can be further evaluated.

The performance statistics obtained with the Lav18 models are similar or worse than the proposed model, namely, over the three experiments $r \geq 0.71$, $\rho \geq 0.76$, MeanErr $\leq 2.7$ dB and MaxErr $= 4$ dB; as opposed to $r \geq 0.97$, $\rho \geq 0.90$, MeanErr $\leq 1.1$ dB and MaxErr $= 3.1$ dB. The main reason is the definition of the internal noise by Lavandier et al. (2018) that leads to two main issues: (1) an underestimation of the effect of HL on better-ear listening for the experiment of Rana and Buchholz (2016) and Rana and Buchholz (2018b) at least at a noise level of 60 dB SPL; (2) an overprediction of the effect of audibility close to the hearing thresholds (Rana and Buchholz, 2018a). In other words, the issue (1) likely means that the level of the internal noise is not high enough to limit the better-ear SNR for HI listeners. Making this level dependent on the external sound level solves this issue. The issue (2) is solved by splitting the audiograms into proportions that resembles OHC and IHC losses, which affects both the predictions at low sensation levels and the effect of increasing audibility. At low sensation level, when the internal noise level is equivalent to $\eta T(n) + B$, the difference between NH and HI listeners is reduced because only 0.7 of the pure-tone audiogram is taken into account. This decreases substantially more the level of the internal noise of HI listeners compared to NH listeners (because of the higher hearing thresholds of HI listeners), so that the predictions are similar at low sensation levels.

The model performance statistics when predicting the experiment of Rana and Buchholz (2018a) are similar. This is due to the parameter that defines the broadband level of the internal noise, which in Lavandier et al. (2018) was optimized separately for each listener group, to predict the difference between conditions within groups and not the difference between groups. Here, the Lav18 models successfully predict the data from Rana and Buchholz (2018a) since there is no significant difference in SRTs between groups. However, the predictions are worse for the experiments from Rana and Buchholz (2016, 2018b) because there is a difference in SRTs between both groups that the Lav18 models cannot predict.

B. Further exploration of the internal noise implementation

Here, a more detailed analysis is provided to illustrate the effect of the external noise level and HL on the internal noise described in Eq. (1), and the expected impact on SI. For this analysis, the average of all the HI listeners’ audiograms across the three main experiments is applied as an example HL. The values after being rounded to the closest multiple of 5 are 15, 20, 25, 35, 45, 50, 60, 60 dB HL for the frequencies 250, 500, 1000, 2000, 3000, 4000, 6000, 8000 Hz.

Figure 7 shows the internal noise spectra for NH and the example HI listener at three different external noise levels (40, 60, and 80 dB SPL). It can be seen that the spectral shape of the internal noise for the NH listener is independent of the external noise level and solely determined by their NH audiogram in dB SPL (i.e., 0 dB HL). The overall internal noise level is identical for the two lowest external noise levels, but then increases substantially at the highest level due to the emerging contribution of the IHC loss component described by Eq. (3). The same behaviour can be observed for the HI listener, except that due to the sloping, high-frequency HL the overall internal noise level is much higher than for the NH listener and increases with increasing frequency. Note, that the effect of hearing aid amplification was disregarded here, which would increase the external masker level and thus the internal noise level. Considering the average HL and the masker stimuli used in the three main experiments, linear amplification according to NAL-RP would result in an increase of the external noise (broadband) level of about 5 dB.

Figure 8 shows the internal noise level for a NH and the example HI listener (see above) as a function of external noise level for the model center frequencies of 516, 2017, and 3937 Hz. For all frequencies and both listeners, the internal noise level is constant at low external noise levels and solely determined by the OHC loss component described by Eq. (2). Above a certain external noise level, the internal noise level starts to increase, due to the emerging contribution of the IHC loss component of Eq. (1). Further increasing the external noise level then leads into a dB-for-dB increase of the internal noise level, as is described by Eq. (3). Again, the overall internal noise level is much higher for the example HI listener than the NH listener, and due to the sloping, high-frequency HL, the influence of the IHC loss component also starts at lower external noise levels than for NH listeners, which becomes further pronounced with increasing frequency. The influence of the NAL-RP amplification on the internal noise levels is plotted as dashed lines. The behaviour of the internal noise levels is similar but the curves are shifted to the left along the x axis by about 5 dB, which represents the broadband increase induced by amplification. Hence, the pure-tone audiograms increase twice the internal noise levels when the signal is amplified according to the HL (as the NAL-RP amplification does): “directly” with the terms $(1 - \eta)T(n)$ and $\eta T(n)$ as well as “indirectly” by increasing the external noise level $N'$. 
For the noises applied in the main experiments with a broadband level of 65 dB SPL, which represents the level of a moderately noisy environment (Weisser and Buchholz, 2019), the level in the model filter bands at the three considered frequencies is 53, 35, and 43 dB SPL. Given that for NH listeners the corresponding internal noise levels are all below 6 dB SPL, the effect of the internal noise on predicted SI is negligible. With respect to the present study, the effect of the internal noise is only relevant in the experiment of Rana and Buchholz (2018a) at the softest sensation level of 0 dB SL. This is already different for the example HI listener, who has a mild-to-moderate HL and, at an external noise level of 65 dB SPL, involves internal noise levels of 17, 34, and 46 dB SPL. Considering the level fluctuations in the masker, the internal noise will swamp (or dominate) a significant portion of the masker signal at mid and high frequencies.

C. Is the proposed internal noise concept in line with the literature?

The proposed implementation of the internal noise shares common aspects with other internal noises described in the literature, in particular the one implemented in the monaural SII (ANSI, 1997). The SII is a SNR-based index (between 0 and 1) that predicts SI in noise. It is obtained by

![Graph showing internal noise levels at three given center frequencies of the model as a function of the external noise level.](https://doi.org/10.1121/10.0002660)
integrating the SNRs across frequency bands using a SII weighting as well as a distortion factor that degrades SI when target levels are higher than a reference. The internal noise of the SII is simply realized by the listener’s pure-tone audiograms averaged across ears and added to a reference internal noise spectrum level (the internal noise level for a NH listener with 0 dB HL). However, there is no distinction of the influences of IHC and OHC loss and the external noise level does not affect the internal noise. The similarity between the two models comes from the way the highest level between the internal noise and the external noise is chosen to compute the SNR in each frequency band. The influence of the distortion factor in the SII could be compared to the influence of the external-level-dependent component of the internal noise in our model, because both degrade SI when the external signal level is higher than a reference. However, they differ in two major ways. First, our internal noise relies on the broadband level of the external noise, whereas the SII distortion factor relies on the level of the target speech in each frequency band. Furthermore, the SII distortion factor is applied to the SNR \textit{a posteriori}; while our level-dependent internal noise is directly involved in the computation of the SNR.

Ching \textit{et al.} (1998) proposed to modify the SII in order to account for the deficit of speech understanding in HI listeners at high sound levels that cannot be explained by loss of audibility. They tested nine procedures to conclude that “the sensation level, as well as the sound presentation level, affect a listener’s ability to make use of audible information.” The internal noise proposed here takes into account these two characteristics since the listener’s audiogram and the external noise level are included in the SNR computation.

The binaural SII “EC/SII” model developed by Beutelmann and Brand (2006) and its extension “BSIM” proposed by Beutelmann \textit{et al.} (2010) combines an EC stage with the monaural SNRs at both ears to predict SI. Both models implement internal noises defining their spectrum levels by adding a parameter equal to 4 and 1 dB, respectively, to a spectrum shaped to the listener’s audiograms. In either model, the internal noises impact the model EC-stage, and thus binaural unmasking. The internal noise can influence the coherence of the resulting overall (internal + external) noises at the listener’s ears, thus reducing the efficiency of the EC mechanism. However, the internal noises also impact the monaural paths of the model, because they tend to reduce the monaural SNR in each frequency band. The internal noise implementations in both models do not take into account the external noise level nor do they differentiate between estimated contributions of IHC and OHC loss. In that way, their implementations are closer to the one of Lavandier \textit{et al.} (2018), in which the audiogram is used to spectrally shape the internal noise and a frequency-independent parameter is applied to define its broadband level.

Our current internal noise implementation can decrease the SNR at the better ear, but it never reduces the overall masker coherence. The impact of the internal noise (including HL) on binaural unmasking is currently modelled such that binaural unmasking is not affected at all, until one of the external signals (masker or target) gets quieter than the internal noise, in which case the binaural unmasking advantage is set to zero in the corresponding frequency band and time frame. This very crude implementation might need refinement in the future, and will require testing on data that specifically assess binaural unmasking for NH and HI listeners.

The definitions of the two internal noise components proposed by Bernstein and Trehiotos (2008) and presented in Sec. I match closely the asymptotic behavior of our internal noise: at low external stimulus level, the internal noise level is equal to $\eta I(n) + B$ [see Eq. (2)], which can be considered as an internal noise floor that is independent of the external stimulus. For a higher external stimulus level, the internal noise increases with the external noise level following a dB-for-dB rule [see Eq. (3)].

D. Hypotheses and limitations of the proposed model

Regarding the individual predictions shown in Sec. III A 4, it is not surprising that the model cannot fully explain the variance seen in the individual SRTs, because it does not consider all aspects of HL. For instance, even though the model takes into account SRM, it is only based on the audiograms and does not consider other aspects of HL such as auditory filter broadening (Glasberg and Moore, 1986) or loss of ITD sensitivity (King \textit{et al.}, 2014; Neher \textit{et al.}, 2011). Moreover, some variation in the data might be influenced by cognitive effects (Neher \textit{et al.}, 2012), motivation or speech material effects (e.g., differences in intrinsic intelligibility of the sentence lists), that are not considered in the model. Given that the NH listeners who participated in all considered experiments were much younger than the HI listeners, it is also likely that age may have contributed to some of the differences seen between groups (Füllgrabe \textit{et al.}, 2015; Neher \textit{et al.}, 2012; Schneider and Pichora-Fuller, 2001), and thus, to the differences seen between data and model predictions. However, since HL naturally increases with age, it is difficult to disentangle the effects of aging and HL.

With reference to the experiments from Rana and Buchholz (2016, 2018a), the proposed model tends to slightly underestimate the difference in SRTs between NH and HI listeners when VS maskers are applied, which indicates that the model slightly underestimates the advantage associated with modulations in the masker envelope for the NH listeners and/or overestimates this advantage for the HI listeners by about 2 dB. This might be explained by the fact that the proposed model does not take into account the larger gap detection thresholds (or reduced temporal resolution) that are typically observed in HI listeners [e.g., Fitzgibbons and Wightman (1982)], which can degrade the benefit associated with listening in the masker’s gaps.

Even though the current model can successfully describe the data from three experiments, its binaural unmasking component was tested only in the “Natural” configuration of Rana and Buchholz (2018b). Existing studies
have shown that an increase of HL leads to lower ITD sensitivity [e.g., Füllgrabe and Moore (2017) and Santurette and Dau (2012)]. Furthermore, a significant negative correlation between ITD sensitivity and SRT was reported by Strelcyk and Dau (2009) and Neher et al. (2011), but in only three test conditions, which might not be enough to draw general conclusions. Our current implementation of the effect of HL on binaural unmasking is very simplistic, which solely considers the effect of reduced audibility but not the effect of reduced ITD sensitivity. This might prove insufficient to fully predict binaural unmasking in HI listeners when more relevant conditions are tested. In addition, as discussed in Sec. IV C, the internal noise implemented in the binaural model proposed by Beutelmann and Brand (2006) or Beutelmann et al. (2010) and the results of Bernstein and Trahiotis (2008) suggest that the coherence of the signal resulting from the combination of the external and internal noises could be lower than the external noise coherence, thus impacting the efficiency of binaural unmasking. Hence, the proposed model should be further tested on datasets that are more relevant for binaural unmasking, including differences in ITDs between the target and masker signals as well as varying the masker coherence at the listener’s ears (Lavandier and Culling, 2010), but measured for HI listeners.

Future studies should also test the proposed model on additional acoustic conditions [such as Beutelmann et al. (2010) did] to further evaluate its general applicability to predict SI in HI listeners. For instance, only energetic noise maskers have been tested so far because the model cannot take into account the informational masking that can occur with speech maskers [e.g., Kidd, Jr. and Colburn (2017)]. Furthermore, it would be relevant to consider reverberant conditions, because reverberation degrades SI in multiple ways (Lavandier and Culling, 2008). It impairs better-ear listening by filling the masker’s gaps (Collin and Lavandier, 2013), and binaural unmasking by decreasing the interaural coherence of the masker (Lavandier et al., 2012). These two effects were well predicted by the NH version of the model, but they were not specifically tested for HI listeners. At high levels, reverberation can also be detrimental to the intrinsic intelligibility of the target, even in the absence of a masker. This effect was successfully predicted by a modified version of the NH model (Leclère et al., 2015), but again, it remains to be tested how well combining this previous model with our internal noise implementation predicts the effects of reverberation for HI listeners. Finally, the non-linear processing present in hearing aids, such as spectral subtraction or non-linear amplification, have not been involved in the present study. The model will need to take into account the influence of this processing to predict SI for HI listeners wearing their hearing aids. In this regard, the general effect of amplification on the internal noise of the model needs to be investigated. Currently, amplification does not only increase audibility of the input signals, but since it increases the overall masker level it also increases the internal noise level of the IHC loss-related component (see Sec. IV B).

Hence, testing the model on a dataset involving conditions with and without amplification would be informative on the effect of amplification on internal noise level.

With respect to the internal noise implementation of the proposed model, the underlying assumptions described in Sec. II B may need to be revised in the future. The first assumption was that the broadband SNR is below 0dB, so that the external level can be approximated by the masker level to define the internal noise. However, for the experiment of Rana and Buchholz (2018a), the SNRs for the 0 dB SL condition are above 0dB. In that case the assumption is violated, even though the average target level was probably not high enough to affect significantly the level of the listener’s internal noise, so that in practice, this violation did not impair the predictions. In addition, as shown by Smeds et al. (2015), daily situations mostly involve positive SNRs, hence, this assumption must be reconsidered to predict intelligibility in real life situations. Similarly, the assumption that the internal noise is solely based on the broadband level of the external noise, may need to be revised. The distortion factor applied in the SII to limit SI at high speech levels, for example, does not only depend on the broadband level of the target speech, but also on its spectrum. Within the proposed model, such frequency dependence of the internal noise might need to be considered in the future.

Another assumption concerns the parameter $\eta$ (equal to 0.7) that is used to divide the listener’s HL into proportions $\eta$ and $1-\eta$, interpreted as a rough estimate of OHC and IHC loss (Sec. II B). Even though this value is close to the value used by Bruce et al. (2013) and Scheidiger et al. (2018), who attribute two thirds of the HL to OHC loss to predict monaural SI, other studies have shown that the proportion of IHC and OHC loss varies across listeners and place (or frequency) on the Basilar membrane [e.g., Moore and Glasberg (2004) and Pieper et al. (2018)]. In this regard, the proposed implementation of a single-value $\eta$ was an intentional simplification, with the aim of limiting the number of fitting parameters and the complexity of the model. However, this may need to be revised in the future.

Finally, the model was both optimized and evaluated on the same dataset involving NH listeners and HI listeners with mild to moderate-severe HL. Even though the model has only a small number of fitting parameters when compared to the number of tested data points, it still needs to be verified on data that is not used to define its parameters. Moreover, the model needs to be tested for more severely impaired HI listeners and listeners with asymmetric HL to evaluate its relevance to predict SI for arbitrary listeners.

V. SUMMARY

A binaural model is proposed that uses the listener’s audiogram to predict the effects of HL and presentation level on SI in noise for HI and NH listeners. This was done by splitting the audiogram into proportions interpreted as rough estimates of OHC and IHC loss and highlighting that the internal noise consists of two components, one related to
elevated thresholds and the other considering suprathreshold effects that depend on the external noise level. The resulting model shows similar predictions to its previous model versions when considering data measured only for NH listeners, and provides accurate results when predicting datasets on which for NH and HI listeners it has been optimized. These involve experimental designs that aimed to evaluate the effects of audibility, spatial configuration, and noise types on SI. Across the five experiments considered in the study, the model predictions are accurate as quantified by the Pearson’s correlation coefficient \( r \) between data and predictions greater or equal to 0.93, a mean absolute prediction error not exceeding 1.1 dB and a maximal absolute error equal to 3.1 dB. The influence of HL on binaural unmasking needs to be further investigated and could lead to future revisions of the model.

**ACKNOWLEDGMENTS**

The authors would like to thank two anonymous reviewers for their helpful comments. The Ph.D. of Thibault Vicente was funded by la Fondation pour l’Audition and Macquarie University. This work was performed within the Labex CeLyA (Grant No. ANR-10-LABX-0060/ANR-16-IDEX-0005).

1. It should be noted that they drew these conclusions using a fixed external noise level (within a condition) and varying the level of the tone such that the external noise level determined the overall stimulus level.

2. When NH and HI listeners are involved, the SRTs are averaged across listeners within each group and then averaged across conditions and groups, giving the same weighting to each group regardless of their number of listeners. The same approach is used for the binaural ratios when removing their mean.

3. The long-term broadband level is preferred in order to provide an internal noise with the same spectral shape as the audiograms. This would not have been possible if the external noise level per band had been considered.

4. The exact numbers of sentences used to model the experiments of Collin and Lavandier (2013), Lavandier et al. (2012), and Rana and Buchholz (2016, 2018a,b) are 60, 120 (for each type of target), 120, 128, and 128, respectively.

5. When using the Lav12 and Col13 models, the input signals are calibrated to the same level, which does not need to be the actual level used in the experiment because the output of these models is independent of the absolute level of their inputs.

6. The actual value of \( \eta \) can be lower than 0.7 because the estimated OHC loss is limited to 57.6 dB. Any listener who presents a loss above 82 dB HL (57.6/0.7) of hearing threshold would have an estimated proportion of OHC loss below 0.7.

7. See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0002660 for plots comparing the predictions of the two models.

8. The value of 4 dB in the version of Beutelmann and Brand (2006) was taken from the literature, while the value of 1 dB was chosen by Beutelmann et al. (2010) because it provided the best correlation between data and predictions in a reference experiment.
ISO 389-2 (1994). *Acoustics—Reference Zero for the Calibration of Audiometric Equipment—Part 2: Reference Equivalent Threshold Sound Pressure Levels for Pure Tones and Insert Earphones (International Organization for Standardization, Geneva, Switzerland, 2001).

Jørgensen, S., Ewert, S. D., and Dau, T. (2011). “Revision and validation of a binaural model for speech intelligibility in noise,” *Hear. Res.* 275(1), 96–104.

Jørgensen, S., Ewert, S. D., and Dau, T. (2013). “A multi-resolution envelope-based model for speech intelligibility,” *J. Acoust. Soc. Am.* 134(1), 436–446.

Kidd, G., Jr., and Colburn, H. (2008). “Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing,” *J. Acoust. Soc. Am.* 125(5), 3328.

Lavandier, M., Jelfs, S., Culling, J. F., Watkins, A. J., Raimond, A. P., and Lavandier, M. (2010). “Prediction of binaural speech intelligibility against noise in rooms,” *J. Acoust. Soc. Am.* 127(1), 387–399.

Lavandier, M., Jelfs, S., Culling, J. F., Raimond, A. P., and Makin, S. J. (2012). “Binaural prediction of speech intelligibility in reverberant rooms with multiple noise sources,” *J. Acoust. Soc. Am.* 131(1), 218–231.

Leclère, T., Lavandier, M., and Culling, J. F. (2015). “Speech intelligibility prediction in reverberation: Towards an integrated model of speech transmission, spatial unmasking, and binaural dereverberation,” *J. Acoust. Soc. Am.* 137(6), 3335–3345.

Moore, B. C. (2008). “The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people,” *J. Assoc. Res. Otolaryngol.* 9(4), 399–406.

Moore, B. C. J., and Glasberg, B. R. (2004). “A revised model of loudness perception applied to cochlear hearing loss,” *Hear. Res.* 188, 70–88.

Neher, T., Laugesen, S., Søgaard Jensen, N., and Kragelund, L. (2011). “Can basic auditory and cognitive measures predict hearing-impaired listeners’ localization and spatial speech recognition abilities?,” *J. Acoust. Soc. Am.* 130(3), 1542–1558.

Neher, T., Lunner, T., Hopkins, K., and Moore, B. C. J. (2012). “Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners (L),” *J. Acoust. Soc. Am.* 131(4), 2561–2564.

Pieper, I., Mauermann, M., Oetting, D., Kollmeier, B., and Ewert, S. D. (2018). “Physiologically motivated individual loudness model for normal hearing and hearing-impaired listeners,” *J. Acoust. Soc. Am.* 144(2), 917–930.

Rana, B., and Buchholz, J. M. (2016). “Better-ear glimpsing at low frequencies in normal-hearing and hearing-impaired listeners,” *J. Acoust. Soc. Am.* 140(2), 1192–1205.

Rana, B., and Buchholz, J. M. (2018a). “Effect of audibility on better-ear glimpsing as a function of frequency in normal-hearing and hearing-impaired listeners,” *J. Acoust. Soc. Am.* 143(4), 2195–2206.

Rana, B., and Buchholz, J. M. (2018b). “Effect of improving audibility on better-ear glimpsing using non-linear amplification,” *J. Acoust. Soc. Am.* 144(6), 3465–3474.

Relato-Iborra, H., Zaar, I., and Dau, T. (2019). “A speech-based computational auditory signal processing and perception model,” *J. Acoust. Soc. Am.* 146(5), 3306–3317.

Rhebergen, K. S., and 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,” *J. Acoust. Soc. Am.* 117(4), 2181–2192.

Santurette, S., and Dau, T. (2012). “Relating binaural pitch perception to the individual listener’s auditory profile,” *J. Acoust. Soc. Am.* 131(4), 2968–2986.

Scheidiger, C., Carney, L. H., Dau, T., and Zaar, J. (2018). “Predicting speech intelligibility based on across-frequency contrast in simulated auditory-nerve fluctuations,” *Acta Acust. Acust.* 104(5), 914–917.

Schneider, B. A., and Pichora-Fuller, M. K. (2001). “Age-related changes in temporal processing: Implications for speech perception,” *Sem. Hear.* 22(3), 227–238.

Smeds, K., Wolters, F., and Rung, M. (2015). “Estimation of signal-to-noise ratios in realistic sound scenarios,” J. Am. Acad. Audiol. 26, 183–196, https://www.thieme-connect.de/products/ejournals/abstract/10.3766/jaaa.26.27.

Strelets, O., and Dau, T. (2009). “Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing,” *J. Acoust. Soc. Am.* 125(5), 3328.

Taal, C. H., Hendriks, R. C., Heusdens, R., and Jensen, J. (2011). “An algorithm for intelligibility prediction of time-frequency weighted noisy speech,” *IEEE Trans. Audio Speech Lang. Process.* 19, 2125–2136.

Vicente, T., and Lavandier, M. (2020). “Further validation of a binaural model predicting speech intelligibility against envelope-modulated noises,” *Hear. Res.* 390, 107937.

Wan, R., Durlach, N. L., and Colburn, H. S. (2014). “Application of a short-time version of the Equalization-Cancellation model to speech intelligibility experiments with speech masks,” *J. Acoust. Soc. Am.* 136(2), 768–776.

Weisser, A., and Buchholz, J. M. (2019). “Conversational speech levels and signal-to-noise ratios in realistic acoustic conditions,” *J. Acoust. Soc. Am.* 145(1), 349–360.