



HAL
open science

Further evaluation of the binaural loudness ratio to predict speech intelligibility in vehicles and rooms

Nikolina Samardzic, Mathieu Lavandier

► To cite this version:

Nikolina Samardzic, Mathieu Lavandier. Further evaluation of the binaural loudness ratio to predict speech intelligibility in vehicles and rooms. *Applied Acoustics*, 2023, 203, pp.109140. 10.1016/j.apacoust.2022.109140 . hal-04284262

HAL Id: hal-04284262

<https://hal.science/hal-04284262v1>

Submitted on 10 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Further evaluation of the binaural loudness ratio to predict speech intelligibility in vehicles and rooms

Nikolina Samardzic¹ and Mathieu Lavandier²

¹ Lawrence Technological University, Department of Engineering Technology, Southfield, MI, 48075, USA

² Univ Lyon, ENTPE, Ecole Centrale de Lyon, CNRS, LTDS, UMR5513, 69518 Vaulx-en-Velin, France

ABSTRACT

This study compared the predictions of speech reception threshold (SRT) across conditions involving spatially distributed speech and noise in a vehicle and in a room, obtained with two prediction methods: the binaural speech-to-noise loudness ratio (BLR) metric proposed by Samardzic and Moore (2021) and the binaural speech intelligibility model (LAV) proposed by Lavandier et al. (2012) used here as a reference. Both methods provide SRT predictions with similar good accuracy for the vehicle experiments that involved background noise generated in a vehicle dynamometer test chamber and on road. The BLR model does not provide accurate SRT predictions for the room experiments that involved both single and multiple stationary noise sources simulated at different positions, for which the LAV model gives accurate predictions. Further analyses indicated that monaural information at the best ear is sufficient to describe the data well for the experiment with background noise generated in a vehicle dynamometer. They also revealed that such monaural analysis is not sufficient for accurate SRT predictions for the experiment with background noise generated on road, nor in the room conditions. Overall, the results indicated that the BLR method does not account for the better-ear listening and binaural unmasking effects underlying the corresponding variations in SRT, particularly in the room experiments.

I INTRODUCTION

The prediction of speech intelligibility is challenging when the speech and background noise at the two ears are different. This is the case when speech is masked with interfering sources in rooms, when understanding speech in noisy backgrounds is also impaired by reverberation (sound reflections in the room). The “cocktail-party effect” (Cherry, 1953) occurs in a variety of challenging listening situations, such as classrooms and office spaces, where a listener tries to understand target speech among other competing sounds. Another common challenging communication situation involves talker and listener inside an operating vehicle. Many individuals have experienced difficulty understanding or following a conversation while driving in a motor vehicle, regardless of age or hearing ability (Meston et al, 2011). Vehicle interior noise, the main cause of the reduction in speech intelligibility, is a sum of partial pressures caused by all air-borne and structure-borne paths from sources of sound and vibration, including input forces from the wind and road, powertrain, intake and exhaust systems. The magnitudes, directions and locations of input forces depend on the vehicle operating conditions and the vehicle design.

Traditional monaural models, such as the Articulation Index (AI; Beranek, 1947), the Speech Intelligibility Index (SII; ANSI S3.5-1997), and the Speech Transmission Index (STI; IEC 60268-16:2003), can result in inaccurate predictions of speech recognition performance both in rooms (Plomp, 1976; Beutelmann and Brand, 2006; van Wijngaarden and Drullman, 2008) and in vehicles (Samardzic and Moore, 2021), because binaural listening is not taken into account. Lavandier and Best (2020) provided a detailed review of different binaural speech intelligibility models. Among those models, Lavandier et al. (2012) proposed a binaural speech intelligibility (LAV) model that resulted in accurate predictions in the measured differences of speech reception thresholds (SRTs, the signal-to-noise ratio, SNR, required to reach 50% intelligibility), in realistic

listening conditions, including a single and multiple noise sources in different rooms. The LAV model incorporates the binaural hearing phenomena called binaural unmasking (BU) and better-ear listening (BE). The differences in the sounds at the two ears help with identifying speech, or, in general, sound, in noise. These differences, including interaural time differences (ITD, resulting in BU) and interaural level differences (ILD, resulting in BE), are due to the sounds coming from different directions from the listener. BE corresponds to the fact that listeners can simply use the information coming from whichever ear offers the better SNR (at a given moment, in a given frequency band). Binaural unmasking corresponds to the release from masking provided by any difference in ITD between the target and masker signals that allows the central auditory system to cancel part of the masker, as per the equalization-cancellation (E-C) theory (Durlach, 1972), thus improving the internal SNR. The LAV model has been validated extensively in anechoic laboratory settings (Jelfs et al., 2011; Culling et al., 2013), as well as in realistic reverberant conditions (Lavandier et al. 2012), considering SRT datasets measured using different procedures and languages (Dutch, English, German) in different laboratories. It was selected here as a representative of the many binaural models that relies on the E-C theory (Lavandier and Best, 2020). The current study extends the evaluation of the LAV model to speech understanding in vehicles.

Previous efforts in modeling speech understanding in vehicles that considered binaural hearing have been limited. Samardzic and Moore (2021) described a model for predicting the speech intelligibility in the steady background noise of a vehicle based on the Binaural Loudness Ratio (BLR). The BLR model is based on the idea that the intelligibility of speech in steady noise is monotonically related to, and can be predicted from, the ratio of the binaural loudness of the speech to the binaural loudness of the noise, both specified in sones. Binaural loudness is the

loudness of sound when heard with both ears. The BLR model provided accurate predictions of the SRTs measured under a variety of conditions, in a vehicle dynamometer and on-road, and for different positions of the talker and listener in the vehicle (Samardzic and Moore, 2021). This accuracy was then evaluated by measuring the standard deviation in the predicted SRT as the change in SNR required to get to the average BLR across conditions, assuming that a given value of BLR corresponds to a fixed level of intelligibility (50% in the case of the SRT). This standard deviation was 1.3 dB, which was smaller than estimated using the SII and STI (2.0 dB and 2.1 dB, respectively). Despite of this first success, the BLR model has never been applied to scenarios beyond speech understanding in vehicles. In the present study, the BLR model was further evaluated in different spatial configurations in rooms, previously shown to involve strong effects of binaural listening (BE and BU). The BLR predictions were systematically compared to the predictions of the LAV model, used here as a reference model, as previously done by other researchers developing binaural intelligibility models (e.g. Andersen et al., 2016). The current study evaluates the BLR model, previously tested only on the vehicle data, to additional room conditions, and systematically compares it to an established binaural intelligibility model (LAV), which in the meantime was extended to vehicle conditions. The present analysis utilizes the speech and noise signals and the SRT measurements from four experiments: two in a vehicle (Samardzic et al, 2012, and Samardzic and Novak, 2013) and two in a room (Lavandier et al, 2012). The speech and noise signals used from Samardzic et al. (2012) and Samardzic and Novak (2013) are based on the SRT measurements within a driving simulator. Prior to being incorporated into a driving simulation, the noise signals were originally recorded in vehicle dynamometer and road driving conditions, respectively. The speech and noise signals used from Lavandier et al. (2012) involve a single or multiple stationary noise sources in different locations in a room. These last experiments

were considered to: 1) test the BLR model in conditions with strong binaural effects, 2) evaluate separately the predictions of binaural masking and better-ear listening, and 3) maintain real-life model application, extending it (from vehicle application) to rooms, by including real-room reverberation and multiple interferers.

II GENERAL METHODS

A. BINAURAL SPEECH INTELLIGIBILITY PREDICTION METHODS

1. BLR: Binaural Speech-to-Noise Loudness Ratio

The BLR is based on the ratio of the binaural loudness of the speech to the binaural loudness of the noise, both specified in sones. The loudness of the speech and noise are calculated separately using the procedure specified in ISO 532-2, as implemented in the “Connect” software of Brüel&Kjær (Nærum, Denmark). The proposed method has the following steps: (1) the binaural loudness of the speech is calculated from the long-term average spectrum of the speech at each ear; (2) the binaural loudness of the noise is calculated from the long-term average of the spectrum of the noise at each ear; and (3) the ratio of these two quantities is taken, giving the BLR.

This ISO 532-2 procedure is based on the model described by Moore et al. (1997) but modified to take into account that loudness does not simply sum across ears; rather, a diotic sound is about 1.5 times louder than the same sound presented monaurally (Moore and Glasberg, 2007; Zwicker and Zwicker, 1991). Moore and Glasberg (2007) modeled this finding using the concept of binaural inhibition; namely, the internal representation of the signal at one ear can be reduced (inhibited) by a signal at the other ear (Scharf, 1969). Briefly, the loudness model includes the following stages: (1) calculation of the effective spectrum reaching the cochlea, taking into account the transmission characteristics of the outer and middle ear; (2) calculation of the excitation pattern of the sound reaching the cochlea, which represents the magnitude of the output of the auditory

filters plotted as a function of filter center frequency (Moore, 2012; Glasberg and Moore, 1990); (3) transformation of the frequency scale to the Equivalent Rectangular Bandwidth number (ERB_N -number) scale, which approximates the way that frequency is mapped to place within the cochlea (Glasberg and Moore, 1990); (4) application of a compressive nonlinearity to the excitation magnitude at each center frequency, to simulate the compression that occurs in the cochlea; (5) application of binaural inhibition based on the relative magnitudes of the compressed signals at the two ears at each center frequency; and (6) summation across center frequencies and across ears to give the overall predicted loudness (for more details, see Moore, 2014).

The BLR method was implemented here using the speech and noise signals and the SRT measurements from all four considered experiments: two in vehicle (Samardzic et al, 2012, and Samardzic and Novak, 2013) and two in rooms (Lavandier et al, 2012). The previous study of the BLR method (Samardzic and Moore, 2021) used the speech and noise signals at the SRT as inputs to the BLR calculation, and the validity of the method was then evaluated by quantifying its ability to produce a constant output (corresponding to the constant intelligibility at the SRT). In the present study, to be compared to the LAV model evaluation, the BLR model was evaluated by first equalizing the target sentences at the level of the background noise (0 dB SNR). The stimuli equalized at 0 dB SNR were submitted to the BLR calculation and resulted in different BLR values for the different conditions. The SNR changes required in the signal inputs to produce the average of these BLR values were taken as the predictions of the changes in SRT across conditions.

In order to compare the relative differences between predicted and measured SRTs, the predicted mean SRT was aligned with the measured mean SRT in each experiment. All the predictions were shifted by the same constant value, therefore, the relative difference between the predictions were not affected by the shift. This model evaluation method was previously used by

Lavandier et al. (2012). It was utilized in the present study for the SRT predictions using the BLR method to obtain a valid basis for a comparison of the SRT predictions from the two methods.

2. LAV: Lavandier et al. (2012) model

The LAV model is a slightly updated implementation (*lavandier2022* model in Lavandier et al., 2022) of the model initially proposed by Lavandier and Culling (2010) and further tested using real room acoustics by Lavandier et al. (2012). It uses the speech and noise signals at the ears to quantify the effects of BU and BE separately, assuming their additive contributions, to ultimately calculate an effective target-to-interferer ratio (TIR). The BU component is calculated using an equation evaluating the binaural masking level difference (BMLD) associated with BU (Culling et al., 2005). This equation is a function of the center frequency of the band, the target and interferer interaural phase difference, and the interferer interaural coherence. The BMLD is then weighted according to the speech intelligibility index (SII) weightings (ANSI S3.5, 1997) and summed across frequency to provide the predicted broadband BU advantage, in dB, which is the predicted improvement in SRT provided by BU. The BE component computes the TIRs as a function of frequency at each ear, selects band-by-band the ear for which the ratio is higher, and sums those SII-weighted TIRs across frequency to provide the broadband better-ear TIR, in dB. Effective TIRs are then obtained by simply adding the BU advantages to the better-ear TIRs. Differences in inverted effective TIR correspond to the differences in predicted SRTs. The predicted and the measured mean SRTs are aligned in each experiment, allowing for a comparison of their relative differences across conditions, as previously explained for the BLR method.

B. SRT MEASUREMENTS: STIMULI, LISTENERS, AND PROCEDURE

1. Vehicle Experiments

The two vehicle experiments were conducted in a vehicle dynamometer test chamber (Samardzic et al, 2012) and road driving conditions (Samardzic and Novak, 2013), to obtain the corresponding background noise recordings, both at a low speed (50 kph) and a high speed (100 kph). The binaural background noise recordings were obtained using a listener head and torso simulator (HATS 4100) in the vehicle dynamometer test chamber, and a SoNoScout Binaural Recording and Analysis System (Type 3653) for the on road conditions.

The speech signal was initially recorded in quiet, in a semi-anechoic chamber. The speech recordings utilized a talker and a listener HATS unit (Type 4128 and 4100, respectively), to accurately capture the directionality of the emitted speech and binaurally perceived speech inside an operating vehicle. The speech material was the “hearing in noise test” (HINT), as developed by Nilsson et al. (1994). The HINT sentences were then prerecorded in the vehicle with the simulated talker (HATS) at a variety of positions. Prior to making the recordings, the HINT calibration signal was played in an anechoic chamber to determine the voltages required to generate the sentences at the desired emitting levels in the vehicle. The spectrum levels of the HINT sentences delivered through the headphones to the human listeners in the driving simulator matched those recorded inside the vehicle to within 0.2 dB, as verified using the HATS and Brüel&Kjær NVH Simulator software.

For both the speech and noise measurements, the listener was always located in the front-left (driver's) seat. The talker was located in the passenger front (F), rear left (L), or rear right (R) seat (Figure 1). Both speech and noise recordings were used to create a driving simulation and they were presented through headphones in the driving simulator.

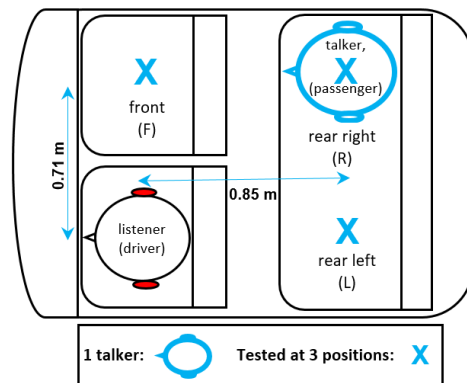


Figure 1: Vehicle speech and noise measurement configurations (Samardzic et al., 2012; Samardzic and Novak, 2013).

The SRT was evaluated using the driving simulation; with 30 normal-hearing (NH) listeners who participated in the dynamometer background noise condition experiment and 9 NH listeners who participated in the on-road background noise condition experiment (4 of whom also participated in the dynamometer experiment). The intelligibility of speech was assessed in the presence of vehicle noise, during the driving simulation. The SRT was measured for both background noise conditions, at 50 kph and 100 kph, and defined as the speech-to-noise ratio required for 50% of sentences to be correctly identified, using the recommended procedure for the HINT.

In the previous BLR study (Samardzic and Moore, 2021), when comparing the BLR to the traditional speech intelligibility metrics, both vehicle dynamometer and road experiments were used for calculating one average BLR value for the entire vehicle measurement dataset, with background noise obtained in a vehicle dynamometer and on road. The background noise was different for these two conditions, and different test subjects participated in the SRT evaluation in each of the two conditions. For these reasons, in the present study, the vehicle dataset was separated into two experiments for analysis; one dataset obtained in the vehicle dynamometer, and one dataset obtained on road. It should be noted that the difference in the number of listeners

between the two vehicle experiments affected the standard error calculations, resulting in a lower standard error for the vehicle dynamometer experiment with a larger number of listeners.

2. Room Experiments

The present study utilizes the room acoustic measurements, specifically the speech and noise signals, from experiments with a single or multiple stationary noise sources in different configurations and distances in a room, as presented by Lavandier et al. (2012) and corresponding to their experiments 1 and 3, respectively.

For the single stationary noise experiment (Figure 2), the noise source was simulated at two distances, 0.65 m (near) and 5 m (far), and three directions from the listener, -25° (left), 0° (front), and 25° (right). The target speech was always at near-right (0.65 m, 25°). Real room listening was simulated by convolving anechoic stimuli with the corresponding binaural room impulse responses (BRIRs) measured in a meeting room. For another set of conditions, spectral-envelope impulse responses (SEIRs) were created by removing the temporal characteristic of the BRIRs while preserving their spectral envelopes. This resulted in the removal of the ITDs required for BU, while preserving the frequency-dependent ILDs needed for BE listening. Therefore, the SEIR stimuli for which BU does not produce any advantage and only BE is involved, allowed for a comparison of the SRTs in the SEIR and BRIR conditions, and quantification of the contribution of BU. This resulted in 12 tested conditions (6 for both BRIR and SEIR conditions).

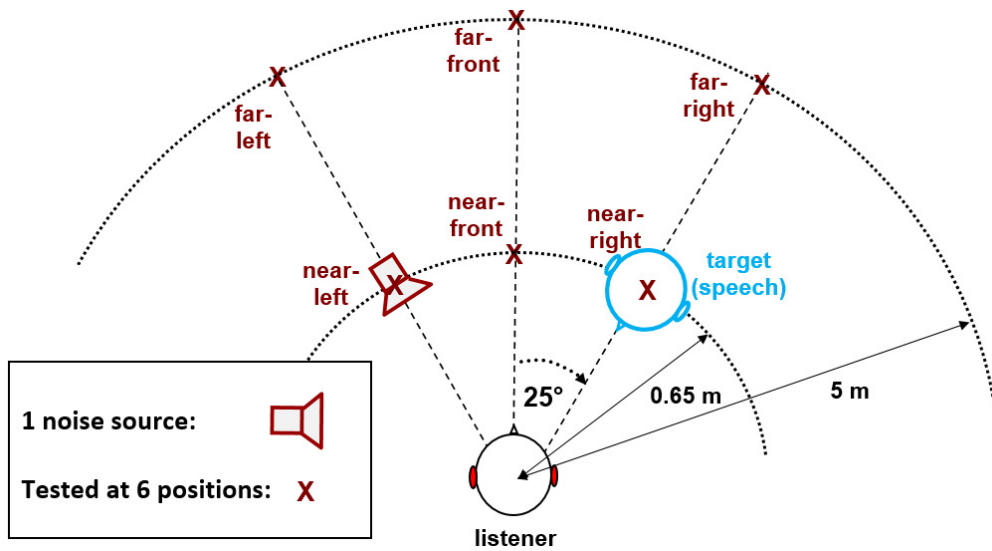


Figure 2: Single stationary noise source measurement configurations (Lavandier et al, 2012).

For the multiple stationary noise experiment (Figure 3), interferers were all tested at either 0.65 m (near) or 5 m (far), in four configurations simulated in the same meeting room as the single stationary noise experiment. For three configurations, the target was always at 0.65 m and 25° (right), where a single interferer was at -25° (left) in the first configuration, a second interferer was added at -25° (left) in the second configuration, and a third interferer was added at 5° (right) in the third configuration. The fourth configuration was bilateral, where the target was at 0.65 m and 0° (front) with one interferer on each side, at -25° and 25°. For this multiple stationary noise experiment, there were 16 tested conditions altogether; 8 BRIR conditions and 8 SEIR conditions.

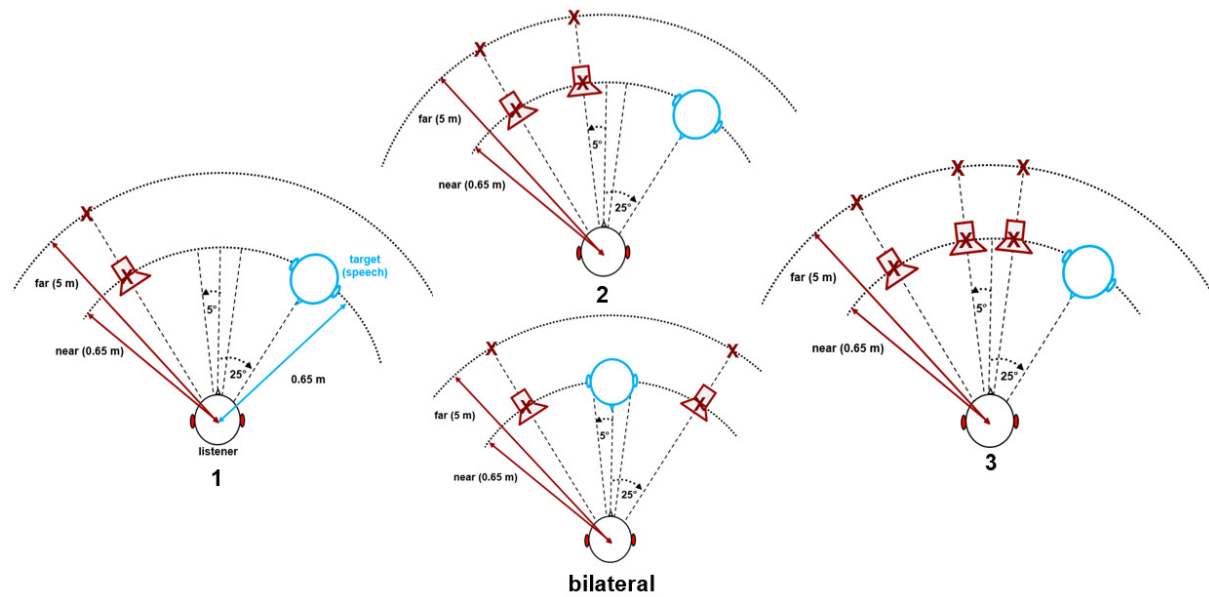


Figure 3: Multiple stationary noise source measurement configurations (Lavandier et al, 2012).

For the single and multiple stationary noise experiments, there were 24 and 32 NH participants, respectively. The speech sentences were from the Harvard Sentence List (IEEE, 1969). SRTs were evaluated with headphones using an adaptive procedure for 50% word correct.

III RESULTS AND DISCUSSION

Figure 4 shows the measured and predicted SRTs for the vehicle experiments, using the LAV and BLR models. Figure 5 shows the data and predictions for the room experiments, also using both models. The main significant effects observed in the data are highlighted in Figure 4 and Figure 5 and are briefly summarized below (for a detailed discussion of these effects and statistical analysis of the data, for Figure 4, see Samardzic et al, 2012, and Samardzic and Novak, 2013, and for Figure 5, see Lavandier et al., 2012.).

In Figure 4, as expected, the SRT increases with increasing vehicle speeds, due to the higher background noise from the vehicle operating at higher speeds. It is generally higher (more difficult to hear) when the talker is directly behind the listener (rear left), compared to when the talker is closer to the listener, in the passenger front, and rear right locations. The difference

between the dynamometer and road speech intelligibility performance is due to the difference in the frequency content of the sources of vehicle interior noise for the two conditions. For example, wind noise, associated with the road driving condition, is a dominant noise source at higher speeds. Different contact forces between the tires and the rolls of the dynamometer and road result in different frequency content of the vehicle body panel vibration and ultimately vehicle interior noise. Additionally, for in-vehicle listening configuration (driver in the front-left location), the “better ear” is the right ear, i.e. closer to the talker and further away from the noisy glass window radiating noise close to the left ear. For the road condition, there are different (additional) airborne and structure-borne sources of noise, such as wind noise, resulting in a different soundscape, compared to the dynamometer condition, and a more symmetrical noise exposure in both ears.

In Figure 5, the difference between BRIRs (black) and SEIRs (grey) indicates the contribution of BU. For the single noise conditions (top panel) and the nearby interferers, the contribution to spatial release from masking (decrease of SRT when the noise is moved to right to left) was larger for BE (SEIR data, grey) than for BU (black-grey difference). Increasing reverberation on the noise source (by increasing its distance to the listener in the room) reduced the influence of an azimuth separation between sources, indicating that head shadow was very limited in the far conditions; but BE then benefited from room coloration that reduced masking for all noise azimuths. BU was still apparent in the far conditions. For the multiple noise conditions (bottom panel), BU remained apparent with increasing number of interferers, whereas the main loss of intelligibility was associated with the loss of BE when interferers were on both sides of the listener (3/near vs 2/near) and on both sides of the target (bilat/near vs 2/near). As for the single noise conditions, increased reverberation reduced BU and head shadow (reduced release from masking between the bilat/far and 2/far conditions), but on average across configurations BE was

not reduced because the loss of head shadow was compensated by the beneficial effect of room coloration.

Objective evaluation of the model performances was accomplished with prediction statistics (Table 1). The Pearson correlation coefficient between measured and predicted SRTs was calculated. The calculation of the mean and maximum prediction errors was obtained by taking the mean/maximum of the absolute difference between predicted and measured SRTs across the tested conditions in each experiment. Standard errors of the measured SRTs were also calculated for each experiment, for comparison with the SRT predictions.

Table 1: SRT prediction statistics for the BLR and LAV models.

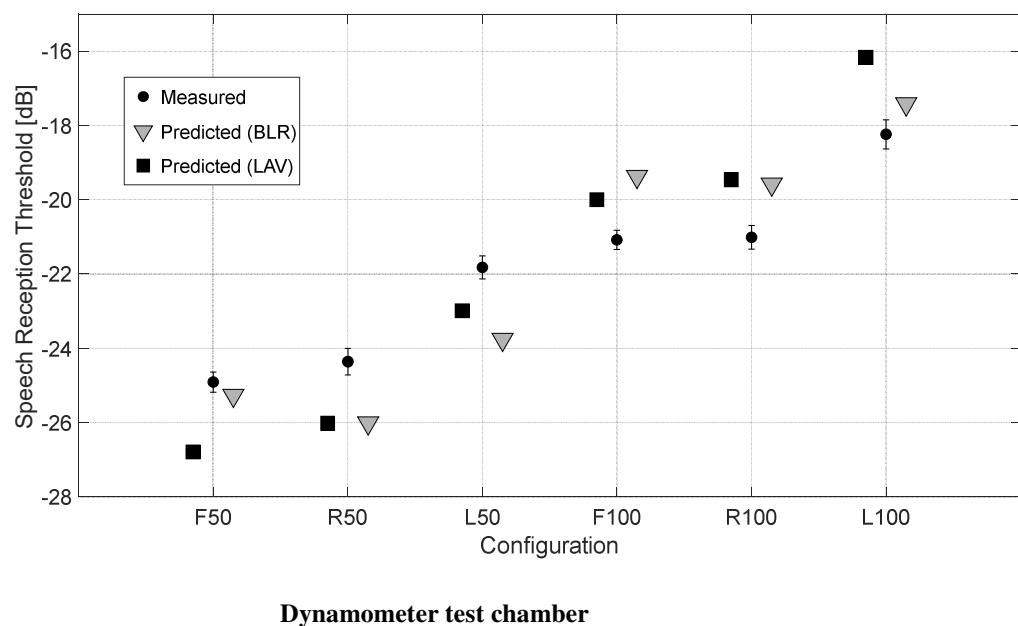
Experiment		Results Figure	Model	Correlation	Error (dB)	
					Mean	Maximum
Vehicle (Figure 1)	Dyno	4 (top)	BLR	0.93	1.3	1.9
			LAV	0.98	1.6	2.1
	Road	4 (bottom)	BLR	0.69	0.7	1.7
			LAV	0.87	0.7	1.1
Room (Figure 2 & 3)	Single Noise Source (Figure 2)	5 (top)	BLR	-0.13	1.4	3.4
			LAV	0.97	0.3	0.6
	Multiple Noise Sources (Figure 3)	5 (bottom)	BLR	-0.26	1.1	2.6
			LAV	0.97	0.3	0.7

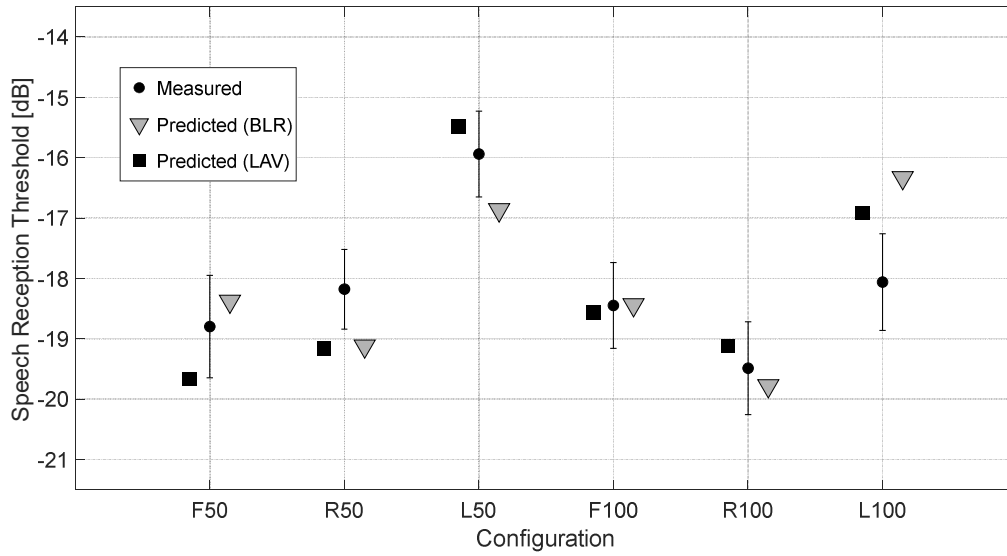
Figure 4 and Table 1 indicate that SRT predictions for the dynamometer vehicle experiment are similar in accuracy for the BLR and LAV models. For the road experiment, the BLR model is similar to the LAV model in terms of prediction errors, however, the correlation is better with the LAV model.

For the dynamometer condition, both models slightly overpredict performance at low speed (50 kph) when the talker is in the front, rear right and rear left locations (F50, R50, L50, respectively). This is possibly due to a listener training effect, resulting in the higher measured SRTs: during the driving simulation used for the SRT measurements, the vehicle listening configurations and speeds were adjusted always in the same sequence (front, rear left, then rear right, at low speed, followed by high speed) for all the listeners within each of the two experiments.

Most of the participants never experienced a driving simulator prior to participating in the study. A degradation in performance in the first three tested conditions for the dynamometer experiment could have occurred as they had to simultaneously drive and listen to the speech signals used for SRT measurements (Wu et al., 2012, 2013). The speech intelligibility models would not capture this effect, thereby overpredicting performance and underpredicting SRTs.

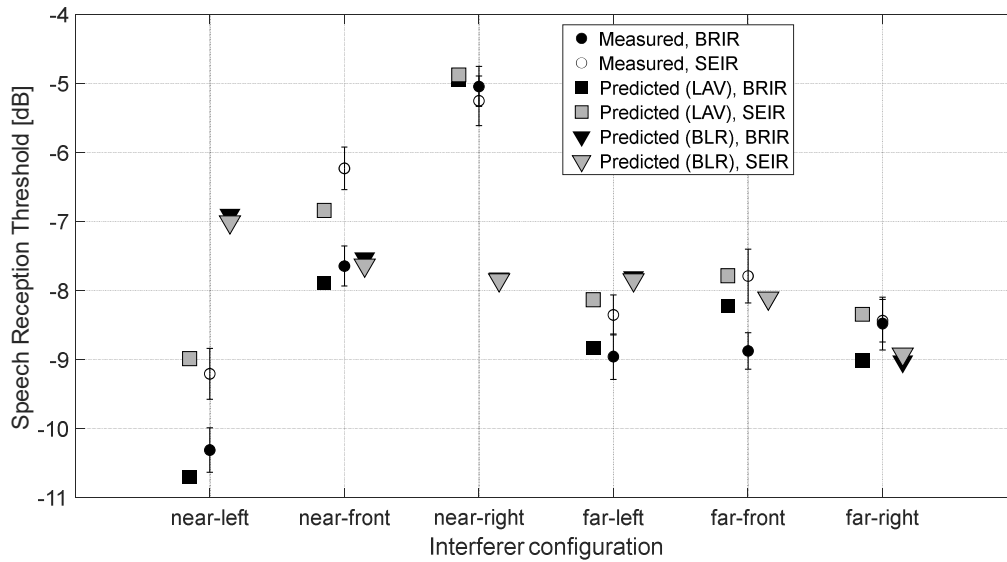
Figure 5 and Table I indicate that the BLR model is unable to predict the SRTs obtained in the room experiments (Lavandier et al., 2012). One factor that contributes to this finding is probably the fact that the BLR model does not take into account the effects of BU and BE. For example, Figure 5 shows similar predictions for BRIR and SEIR using the BLR model, indicating that this model does not predict the advantage provided by BU. The LAV model captures these effects and can describe the SRT measurements well.



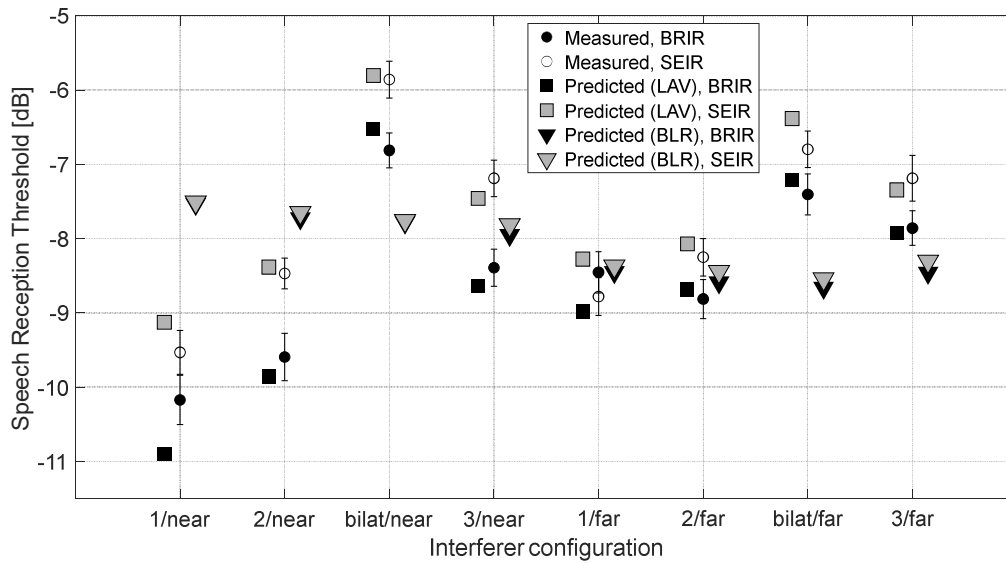


On road

Figure 4: LAV (black) and BLR (grey) SRT predictions of vehicle dynamometer data (top panel; Samardzic et al, 2012) and vehicle road data (bottom panel; Samardzic and Novak, 2013). The listener is always the driver (front left), and the target talker is in the passenger front (F), rear right (R) or rear left (L) positions, for the vehicle at 50 or 100 kph (i.e. for F50, the talker is in the front position of the vehicle at 50 kph). A small horizontal offset has been added to the model predictions to reduce symbol overlap.



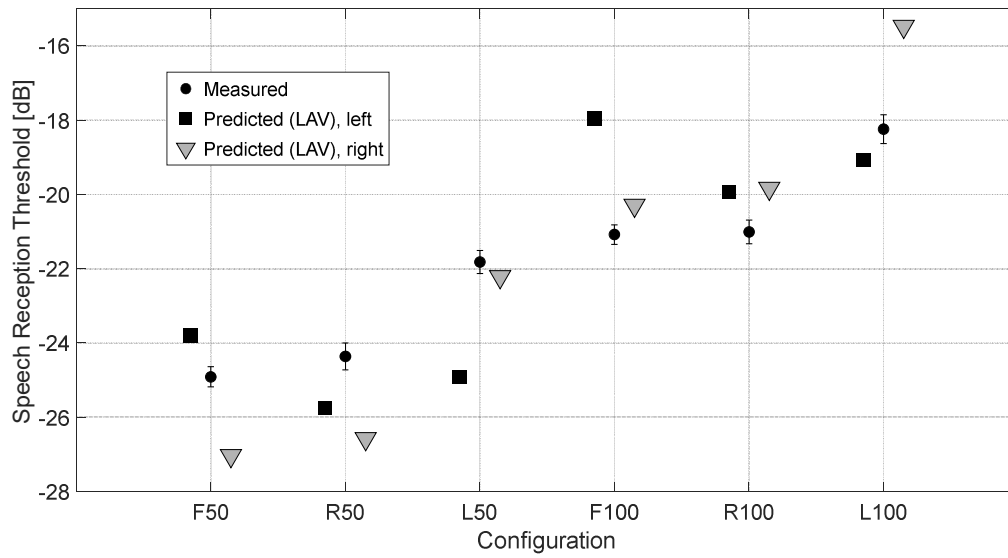
Single stationary noise source



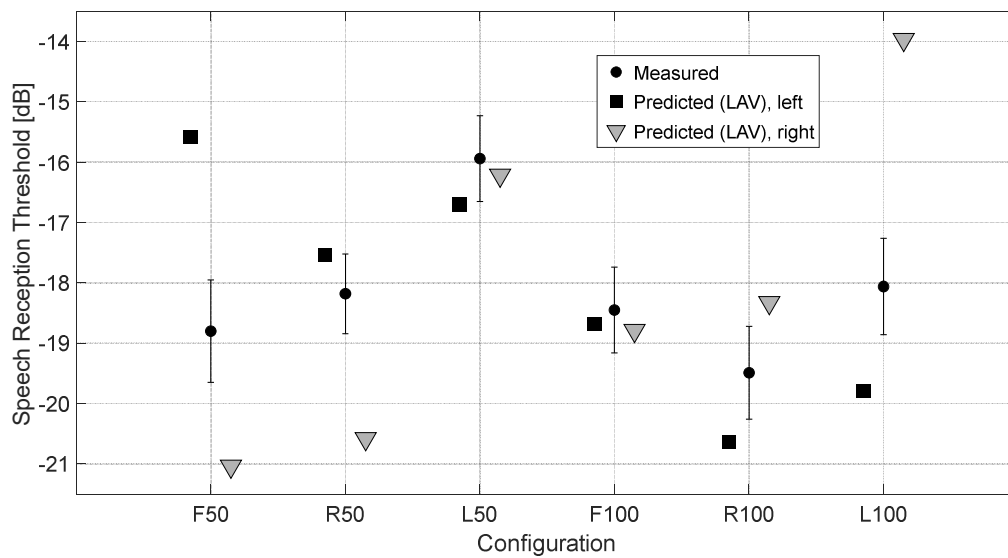
Multiple stationary noise source

Figure 5: LAV (full lines) and BLR (dashed lines) SRT predictions for the room experiments. For the single stationary noise source conditions (top panel), the target speech was always at near-right (0.65 m, 25°), while the noise source was simulated at two distances, 0.65 m (near) and 5 m (far), and three directions from the listener, -25° (left), 0° (front), and 25° (right). For the conditions with multiple stationary noise sources (bottom panel), the target was at 0.65 m, at 25° (1, 2, and 3) or 0° (bilateral/bilat.), while the interferers were simulated at 0.65 and 5 m (near/far). A single interferer was at -25° (1), two interferers were at -25° and -5° (2), three interferers were at -25°, -5°, and 5° (3), or two interferers were at -25° and 25° (bilat). A small horizontal offset has been added to the model predictions to reduce symbol overlap.

An additional analysis with diotic signals was performed, motivated by the fact that, although both models work relatively well for the vehicle conditions, a large difference is observed for the room conditions. In order to investigate the impact of binaural hearing in the different experiments, the SRT predictions using the LAV model were replicated using diotic signals. Either only the left ear signals are used for both ears, or only the right ear signals are used for both ears, to create diotic signal inputs to the model, both for the vehicle (Figures 6) and room (Figures 7) experiments. Table 2 shows SRT prediction statistics using these diotic signals.

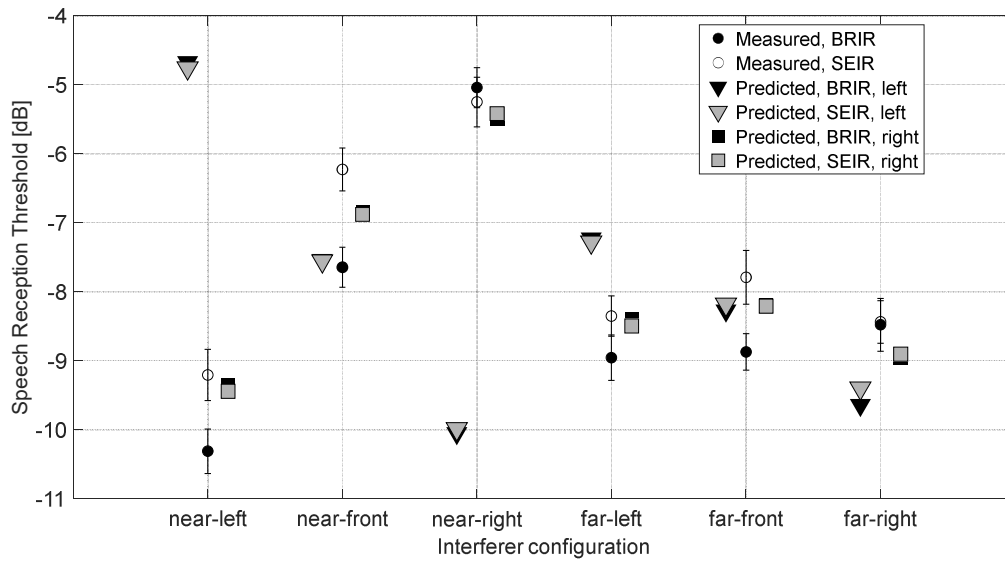


Dynamometer test chamber

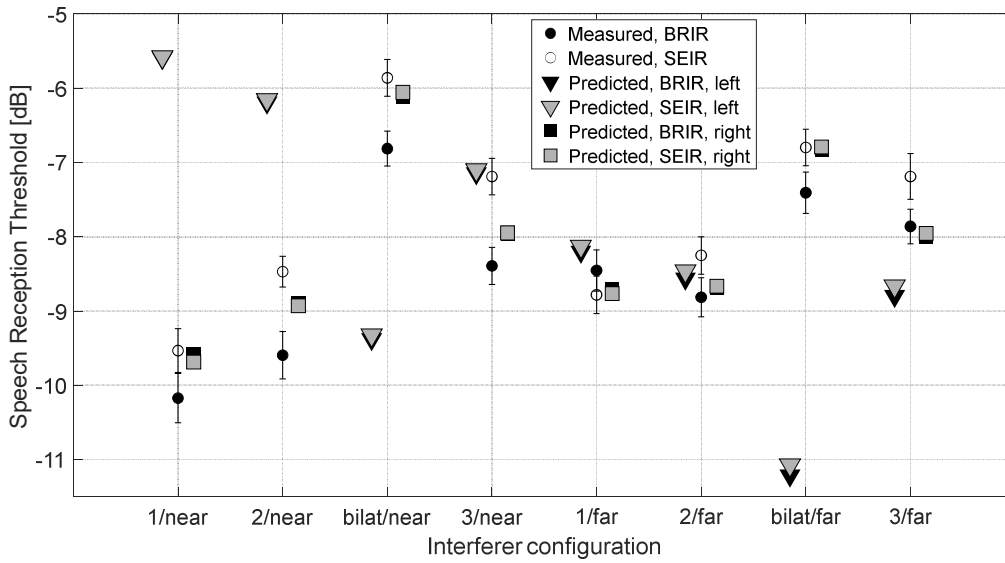


On road

Figure 6: Diotic SRT predictions for the vehicle experiments in a dynamometer test chamber (top panel; Samardzic et al., 2012) and on road (bottom panel; Samardzic and Novak, 2013). A small horizontal offset has been added to the model predictions to reduce symbol overlap.



Single stationary noise source



Multiple stationary noise sources

Figure 7: Diotic SRT predictions for the room experiments with a single stationary noise source (top panel; Exp. 1 from Lavandier et al., 2012) and with multiple stationary noise sources (bottom panel; Exp. 3 of Lavandier et al., 2012). A small horizontal offset has been added to the model predictions to reduce symbol overlap.

Table 2: SRT prediction statistics for diotic signals with the LAV model.

Experiment		Results Figure	Ear Signal	Correlation	Error (dB)	
					Mean	Maximum
Vehicle (Figure 1)	Dyno	6 (top)	Left	0.75	1.8	3.1
			Right	1.00	1.6	2.8
	Road	6 (bottom)	Left	0.42	1.3	3.2
			Right	0.44	1.8	4.1
Room (Figure 2 & 3)	Single Noise Source (Figure 2)	7 (top)	Left	-0.68	2.3	5.6
			Right	0.94	0.5	1.0
	Multiple Noise Sources (Figure 3)	7 (bottom)	Left	-0.74	2.1	4.6
			Right	0.91	0.4	0.8

For the vehicle experiment in the dynamometer test chamber, a diotic version of the speech signal, using the right ear, resulted in a reasonably accurate prediction of the SRTs (Figure 6, top panel). There is a high correlation (0.996, approximated to 1) between measured and predicted SRTs, although the maximum error is slightly higher compared to the SRT prediction using the dichotic version of the signal displayed in Figure 4 (2.8 dB, compared to 2.1 dB). The diotic signal SRT prediction using the left ear shows a lower correlation (0.7) between the measured and predicted SRTs, and a higher mean and maximum errors compared to the right ear diotic SRT predictions. These results indicate that monaural predictions based only on the signals at the right ear are almost sufficient for accurate SRT predictions for this vehicle experiment. This indicates that the binaural effects were minimal in these conditions where the two models were equivalent.

For the vehicle experiment on road, the diotic versions of the signals do not provide accurate SRT predictions (Figure 6, bottom panel). The right ear alone, presumably the better ear, is not sufficient to describe the data. There is a low correlation between the measured and predicted SRTs using both diotic versions of the signal (0.42 and 0.43, for left and right ear, respectively). Both ears' signals are necessary to obtain the most accurate SRT predictions (Figure 4, bottom panel), indicating that BE (which allows to switch between ears to get the better SNR) and/or BU are important to consider in these conditions. This might explain why, in these on road conditions,

the LAV model provide predictions more correlated with the data compared to the BLR model that does not account for these binaural effects.

For the room experiments, the diotic predictions with the right ear signals capture the effect of BE (Figure 7), thus indicating that the right ear was probably the better ear in all conditions. For the single noise conditions (top panel), the right ear should be the better ear. For the multiple noise conditions (bottom panel), the right ear should be the better ear in conditions 1, 2 and 3; it is not the better ear in the bilateral condition, in which there is no better ear because of the symmetric set-up, apart from a potential asymmetry created by room acoustics. So, in practice, having the right ear signals is almost sufficient to capture the effect of BE. The right-ear diotic signals provide slightly less correlated SRT predictions than the dichotic signals for the room experiments with single and multiple noise sources (correlation coefficient of 0.93 and 0.91, compared to 0.97 and 0.97). But more importantly (and as expected), the diotic predictions cannot capture the effect of BU that triggers the difference between the gray (SEIR) and black (BRIR) symbols in Figure 7.

In summary, both models contain a processing stage that resembles peripheral filtering and frequency selectivity, but they account for different aspects of binaural hearing on perception. The BLR model considers the effect of binaural hearing on the perception of loudness, but it does not account for the effects of the phase difference across ears on which BU rely, nor the effects of BE that allows switching across ears to get the best possible SNR to understand the target speech. The present results indicate that speech intelligibility in noise apparently relies more on BE and BU, while perceived loudness could well be more relevant when considering listening effort. The BLR metric produced comparable results to the LAV model in the vehicle conditions involving mostly monaural information (dynamometer test chamber conditions), but then was less and less accurate

when the binaural (BE and BU) effects were more and more involved (vehicle on road conditions, and room experiments).

In the previous BLR study (Samardzic and Moore, 2021), the SRT prediction error was quantified as a change in SNR to get constant BLR across all tested conditions. In theory, if a given intelligibility metric is accurate, the value of that metric at the SRT should be constant across conditions. The standard deviation of the SRT prediction errors was calculated, however, this type of error assessment does not indicate whether the results are systematically overpredicting or underpredicting the SRT. Also, evaluating a model by checking that it produces a constant output (at the SRT) could be very misleading because, for example, a bad “model” that produces a constant value whatever the input signals are would pass the test perfectly even if it cannot predict anything (Lavandier et al., 2022). Instead, in the present study, mean/maximun absolute prediction errors across conditions were calculated as a better measure of the model performance.

It should also be noted that for the room experiments, although the target speech recordings used for the SRT measurements were equalized at 0 dB SNR, the absolute presentation levels (dB SPL), for both target and masker, were not reported. In the present study, the BLR predictions of the Lavandier et al. (2012) data were based on a noise level presentation approximated to 60 dB SPL across ears. Any potential BLR differences that may result from using different (than actual, experimental) noise signal levels are unknown.

IV CONCLUSIONS

The BLR model provided SRT predictions with good and similar accuracy as the LAV model for one of the vehicle experiments, more so in the vehicle dynamometer test chamber conditions than in the on-road conditions. The diotic SRT predictions of the vehicle experiments revealed that monaural predictions at the best ear are sufficient to predict the SRT in background

noise associated with the vehicle dynamometer test chamber, while it was not the case for the road vehicle conditions. The comparison of the diotic and binaural predictions for the room experiments confirmed the importance of the two binaural effects BE and BU on intelligibility. The BLR method does not account for these effects and thus progressively failed to predict SRTs when they were more and more involved (vehicle on road conditions, and room experiments). The BLR model was not able to predict the SRT at the same level of accuracy as the LAV model in three out of the four datasets tested. It worked well in the experiment with less binaural effects. In practice, the LAV model is recommended for situations with strong binaural effects involved.

Acknowledgments

M.L. takes part in the LabEx CeLyA (Grant No. ANR-10-LABX-0060). We would like to thank Yi Shen for helpful comments and suggestions on the paper.

References

- 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, 1908–1920.
- ANSI (2017). ANSI S3.5-1997, “American National Standard Methods for the Calculation of the Speech Intelligibility Index” (American National Standards Institute, New York).
- Beutelmann, R., and Brand, T. (2006). “Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners,” *J. Acoust. Soc. Am.* 120(1), 331–342.
- Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (2005). “Erratum: The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources,” *J. Acoust. Soc. Am.* 118, 552.
- Culling, J.F., Lavandier, M., and Jelfs, S. (2013). Predicting binaural speech intelligibility in architectural acoustics. *The Technology of Binaural Listening*, ed. J. Blauert, 427-447. Berlin-Heidelberg-New York, NY: Springer.
- Durlach, N. I. (1972) “Binaural signal detection: Equalization and cancellation theory,” in *Foundations of Modern Auditory Theory* edited by J. Tobias, volume II (Academic, New York) pp. 371-462.

French, N.R. and Steinberg, J.C. (1947) “Factors governing the intelligibility of speech sounds”, *J. Acoust. Soc. Am.* 19, 90–119.

Glasberg, B.R. and Moore, B.C.J. (1990) “Derivation of auditory filter shapes from notched-noise data”, *Hear. Res.*, 47, 103–138.

H.J.M. Steeneken and T. Houtgast, “A physical method for measuring speech-transmission quality”, *J. Acoust. Soc. Am.*, 67, 318–326, (1980).

IEC (2003). IEC 60268-16: 2003, *Sound System Equipment - Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index* (International Electrotechnical Commission, Geneva, Switzerland).

ISO 532-2:2017 (2017) Acoustics — Methods for calculating loudness - Part 2: Moore- Glasberg method, International Standard, International Organization for Standardization, Geneva, Switzerland.

Jelfs, S., Culling J. F., Lavandier M. (2011), “Revision and validation of a binaural model for speech intelligibility in noise” *Hear. Res.*, 275, 96-104.

Lavandier M. and Best V. (2020) *Modeling binaural speech understanding in complex situations* In J. Blauert & J. Braasch (ed.), *The technology of binaural understanding*, pp. 547-578, Springer Nature Switzerland.

Lavandier, M., Vicente, T., Prud’Homme L. (2022) “A series of speech intelligibility models in the Auditory Modeling Toolbox” * Conditionally accepted for publication in *Acta Acustica*.

Lavandier, M., Jelfs, S., Culling, J.F., Watkins, A.J., 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, 218–231.

Lavandier, M., Culling J. F. (2010), “Prediction of binaural speech intelligibility against noise in rooms”, *J. Acoust. Soc. Am.*, 127 (1), 387-399.

Meston, C.N., Jennings, M.B., Cheesman, M.F. (2011) „Older Adults“ Views of their Communication Difficulties and Needs while Driving in a Motor Vehicle“, *Canadian Journal of Speech-Language Pathology and Audiology*, Vol. 35, No. 4, pp. 312-321.

Methods for the calculation of the speech intelligibility index, American National Standards Institute ANSI S3.5-1997, New York (2017).

Moore, B.C.J. (2012) *An Introduction to the Psychology of Hearing*, 6th edn, Brill, Leiden, The Netherlands.

Moore (2014), “Development and current status of the “Cambridge” loudness models”, *Trends Hear.*, 18, 1–29.

Moore, B.C.J., Glasberg, B.R. and Baer, T. (1997) “A model for the prediction of thresholds, loudness and partial loudness”, *J. Audio Eng. Soc.*, 45, 224–240.

Moore, B.C.J. and Glasberg, B.R. (2007) “Modeling binaural loudness”, *J. Acoust. Soc. Am.*, 121, 1604–1612.

Nilsson, M., Soli, S.D. and Sullivan, J.A. (1994) “Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise”, *J. Acoust. Soc. Am.*, 95, 1085– 1099.

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),” *Acustica* 34, 200–211.

Qatu, M.S., Abdelhamid, M.K., Pang, J. and Sheng, G. (2009) “Overview of Automotive Noise and Vibration”, *Int. J. Vehicle Noise and Vibration*, Vol. 5, pp. 1-35.

Samardzic, N. and B.C.J. Moore (2021) “Binaural speech-to-noise ratio at the speech reception threshold in vehicles”, *Noise Control Eng. J.*, 69, 173–179.

Samardzic, N., Novak, C. and Gaspar, R. (2012) “The evaluation of speech intelligibility in a simulated driving environment using the Hearing in Noise Test (HINT)”, *Int. J. Vehicle Noise. Vib.*, 8, 318–336.

Samardzic, N. and Novak, C. (2013) “The analysis of the reduction in vehicle speech intelligibility for normal hearing and hearing impaired individuals in a simulated driving environment with contributions from the ordered and masking noise sources”, *J. Audio Eng. Soc.*, 61, 676–687.

Scharf, B. (1969) “Dichotic summation of loudness”, *J. Acoust. Soc. Am.*, 45, 1193–1205.

Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision research*, 122, 105-123.

Sound System Equipment — Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index, International Electrotechnical Commission, IEC 60268-16: 2003, Geneva, Switzerland, (2003).

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.

Wu, Y. H., Stangl, E., Bentler, R. A., & Stanziola, R. W. (2013). The effect of hearing aid technologies on listening in an automobile. *Journal of the American Academy of Audiology*, 24(06), 474-485.

Wu, Y. H., Aksan, N., Rizzo, M., Stangl, E., Zhang, X., & Bentler, R. (2014). Measuring listening effort: Driving simulator vs. simple dual-task paradigm. *Ear and hearing*, 35(6), 623.

Zwicker, E. and Zwicker, U.T. (1991) "Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution", *J. Acoust. Soc. Am.*, 89, 756–764.