



HAL
open science

The Utilization of Psychometric Functions to Predict Speech Intelligibility in Vehicles

Nikolina Samardzic, Mathieu Lavandier, Yi Shen

► **To cite this version:**

Nikolina Samardzic, Mathieu Lavandier, Yi Shen. The Utilization of Psychometric Functions to Predict Speech Intelligibility in Vehicles. SAE International Journal of Vehicle Dynamics, Stability, and NVH, 2023, 8 (1), pp.10-08-01-0002. 10.4271/10-08-01-0002 . hal-04818081

HAL Id: hal-04818081

<https://hal.science/hal-04818081v1>

Submitted on 6 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.

The Utilization of Psychometric Functions to Predict Speech Intelligibility in Vehicles

Preprint of

SAMARDZIC N., LAVANDIER M., SHEN Y.

The utilization of psychometric functions to predict speech intelligibility in vehicles

SAE International Journal of Vehicle Dynamics, Stability, and NVH, 8 (1), pp. 21-30, January 2024

Abstract

In this study, a novel assessment approach of in-vehicle speech intelligibility is presented using psychometric curves. Speech recognition performance scores were modeled at an individual listener level, for a set of speech-recognition data previously collected under a variety of in-vehicle listening scenarios. The model coupled an objective metric of binaural speech intelligibility (i.e. the acoustic factors) with a psychometric curve indicating the listener's speech recognition efficiency (i.e. the listener factors). In separate analyses, two objective metrics were used with one designed to capture spatial release from masking and the other designed to capture binaural loudness. The proposed approach is in contrast to the traditional approach of relying on the speech recognition threshold, the speech level at 50% recognition performance averaged across listeners, as the metric for in-vehicle speech intelligibility. Results from the presented analyses suggest the importance of considering speech recognition accuracy across a range of signal-to noise ratios rather than the speech recognition threshold alone, and the importance of considering individual differences among listeners when evaluating in-vehicle speech intelligibility.

1. Introduction

The perceived vehicle sound quality is strongly correlated with psychoacoustic metrics [1] that quantify subjective perception about objective physical characteristics of sounds. The development of the vehicle sound quality performance is becoming more stringent with the increasing use of electric vehicles [2]. Vehicles are viewed as one of the most challenging speech communication environments for both normal-hearing and hearing-impaired listeners [3 – 6]. The effectiveness of in-vehicle speech communication can also affect the perception of the overall vehicle quality. Vehicle interior noise is the main cause of the reduction in speech intelligibility. Previous research investigated in-vehicle speech intelligibility and the effects of the vehicle interior noise spectral content, speech directivity, talker/listener distance, binaural listening, hearing profile [7, 8, 9, 10, 11, 12, 13], and speech optimization for improved intelligibility of voice messages by driving assistance systems in vehicles [14].

The prediction of speech intelligibility is challenging when the speech and background noise at the two ears are different. In such situations, it is crucial to consider binaural speech intelligibility models, as shown in recent investigations of in-vehicle speech intelligibility [15, 16]. Previously, binaural recordings of speech stimuli in various vehicle running conditions have been made and speech recognition data were collected from a cohort of normal-hearing individuals [17]. This raw data (binaural recordings) were used in a recent study [16], where two models of binaural hearing have been utilized to fit the data with some success in predicting the average speech recognition threshold (SRT), i.e. the speech level relative to the noise level (SNR, signal-to-noise ratio) at which a 50% recognition performance is achieved across the listeners.

Modeling in-vehicle speech ineligibility while focusing only on the average SRT has at least two drawbacks. First, this approach would not be able to generate predictions of speech recognition performance in terms of percentage correct in word recognition for a given test condition. In practice, it is often useful to estimate the speech recognition scores for a fixed speech level in various test conditions, rather than the speech level required to reach 50% performance in these conditions. Second, even among listeners with normal-hearing, individual differences in the efficiency of speech recognition may be

significant [18, 19]. These types of individual differences have been neglected in research investigations of in-vehicle speech intelligibility. The aim of the present study was to use predictive models to generate predictions of not only the average performance across listeners, but also the expected variability among individual listeners. Representing the spread of the predicted speech recognition performance could better assist in vehicle design and regulatory decision making.

The current study directly addresses the two limitations stated above by extending binaural speech intelligibility models to make predictions of in-vehicle speech recognition scores for individual listeners. In particular, psychometric functions are developed that link these scores to the objective predictive indices produced by models of speech intelligibility. Fitting psychometric functions is a common technique in psychoacoustics to establish the relationship between the proportion of correct responses as a function of signal strength (e.g., [20, 21, 22]). For example, for a speech intelligibility task, as the sound pressure level of the speech signal increases, listeners' percent correct will increase monotonically from chance to 100% correct [23 – 26]. As mentioned above, the SRT is defined at some point along the psychometric function (e.g., the 50%-correct point). The slope of the psychometric function at the SRT indicates the rate at which speech recognition score changes with SNR [27].

The applications of psychometric function include evaluating speech intelligibility models [28], the effects of the hearing aids' algorithms on speech intelligibility [29, 30, 31], predicting speech intelligibility using automatic speech recognition system and neural networks [32], electrophysiological and behavior measures of speech intelligibility [33], and the effects of aging on the masked speech recognition tasks [34].

The current study formulates psychometric functions that relate objective metrics of speech intelligibility to speech recognition performance and introduces computational procedures to fit the psychometric functions to individual listeners' responses from a speech recognition experiment. The psychometric function threshold and slope parameters could vary from one listener to another, but were assumed independent of stimulus conditions. The effects of stimulus condition and individual listeners were separated into the front-end objective metric and the psychometric function, respectively. Such evaluation is new for in-vehicle speech intelligibility applications.

In the following sections, the formulation of the psychometric functions and the computational steps to fit such function is introduced. This is followed by brief descriptions of the objective metrics of binaural speech intelligibility utilized in the current study. Finally, analysis results by fitting the psychometric functions to a set of previously collected speech recognition data involving background noise and speech obtained in a vehicle dynamometer test chamber are presented.

The present analysis utilizes some of the raw data of Samardzic et al. (2012) [17], specifically the speech and noise spectra at the speech reception threshold (SRT). For a detailed discussion of the significant characteristics of the background noise, speech and background noise measurements, vehicle and human participants' information and jury testing using a driving simulator, see [17]. The present study also uses two speech intelligibility models previously used to predict the corresponding SRTs [15,16] and focuses on a dataset for which they give similarly good predictions. What is new here is the psychometric function method that is proposed both to compare model predictions and allow prediction of full psychometric functions instead of SRT predictions alone. Some data and model predictions at the SRT are re-used here to illustrate this new method; while additional model predictions above and below the SRT were produced to be able to consider full psychometric functions.

2. Predicting Speech Recognition Performance Using Psychometric Functions

Numerous objective metrics of speech intelligibility, such as the Speech Intelligibility Index (SII, ANSI S3.5, 1997, [35]), the Speech Transmission Index (STI, IEC 60268-16, 2003, [36]), and the Short-Time Objective Intelligibility (STOI, Taal et al., 2011, [37]), have been developed over the past decades. These models are developed to capture the influences from various stimulus conditions on speech intelligibility performance. The conditions such as the spectrum of the background noise, the amount of reverberation, and the spatial locations of the speech and noise sources relative to the listener, would yield the same speech recognition performance as long as the objective (model) predictors calculated for the conditions are identical.

In order to predict speech recognition scores for a listener, a psychometric function can be used to relate the objective predictor to proportion correct in speech recognition. The psychometric function takes the form of a logistic function:

$$p_{\text{correct}} = \frac{1}{1+e^{-\beta(z-\alpha)}} , \quad (1)$$

where p_{correct} is the proportion correct in speech recognition; z is the objective metric of speech intelligibility; and α and β are the threshold and slope parameters of the psychometric function, respectively. The threshold parameter α corresponds to the value of the objective metric that corresponds to 50% correct in speech recognition). The slope parameter β specifies, at 50% correct, the rate at which the performance score increases as the value of the objective metric increases. A greater β value indicates a steeper slope, while a smaller β value indicates a shallower

slope. The logistic function, as specified in Eq. 1 is a common way to formulate psychometric function in psychoacoustics [22, 38, 39]. Specifically, the logistic psychometric function is used to describe speech recognition performance as a function of SNR [23, 27, 40, 41].

In the current modeling approach, the threshold and slope parameters of the psychometric function are assumed independent of stimulus conditions but allowed to vary from one listener to another. Therefore, the current model implementations separate the effects of stimulus condition and individual listeners into the front-end objective metric and the psychometric function, respectively.

The psychometric function parameters may be estimated using data collected from a speech recognition experiment. This is done first by computing the objective metrics associated with the stimuli used during all test trials in the experiment. For example, based on the stimuli on the i th trial of the experiment, the model produces a metric z_i . For any pair of α and β parameters, the proportion correct in keyword recognition p_i can be predicted by the logistic psychometric function according to Eq. 1. Let r_i indicate the number of keywords out of the total of m_i keywords correctly recognized by the listener on the i th trial. Then the log likelihood of the model based on the i th trial is:

$$\log l_i(\alpha, \beta) = r_i \log p_i + (m_i - r_i) \log(1 - p_i). \quad (2)$$

If the speech recognition performance is scored on the sentence level, then $m_i = 1$, and r_i is either 1 or 0 indicating the success or failure in recognizing the sentence, respectively. The model can then be fitted to the experimental data by finding the α and β values such that:

$$\{\alpha, \beta\} = \underset{\alpha, \beta}{\operatorname{argmax}} \sum_i \log l_i(\alpha, \beta). \quad (3)$$

In practice, the estimated α and β parameters may be influenced by the test stimuli used in the speech recognition experiment. For example, the threshold and slope of the psychometric function may vary depending on whether the target speech is nonsense syllables, monosyllabic words, spondees, or sentences; the slope of the psychometric function may also change when the background noise contains envelope fluctuations or consists of competing speech (e.g., [27]). On the other hand, when listeners are tested using the same stimuli in an experiment (e.g. same speech and noise material), the variations in the estimated α and β parameters across listeners reflect the listener-specific factors that impact speech understanding. In such a situation, a higher α estimate suggests that the listener requires a greater degree of speech audibility to achieve 50% recognition, while a lower α estimate suggests that the listener is highly efficient in speech recognition. Further, a higher β estimate suggests the listener's speech recognition performance may improve sharply as speech audibility improves, while a lower β estimate suggests the expected benefit from the improvement in speech audibility for the listener may be limited.

3. Applying Psychometric Functions to In-Vehicle Speech Recognition

This section provides a demonstration of the modeling approach described in the previous section. Psychometric functions were fitted to previously collected speech recognition data. First, the participants, stimuli, and procedure involved in the original speech recognition study are briefly described. Two objective metrics that were used to generate predictions of speech intelligibility in terms of SRT are then described. This is followed by a description of the computational procedures for fitting the psychometric functions. Finally, results from the analysis are presented. The goal of this demonstration is to illustrate:

(1) how to link the objective metrics to speech recognition accuracy through fitting the psychometric function, and (2) how the psychometric functions from individual listeners may vary widely even under acoustically identical testing conditions.

3.1. In-vehicle speech recognition data

The intelligibility of speech was assessed in the presence of vehicle background noise, during a driving simulation, using a Brüel&Kjær NVH Desktop Simulator. The SRT was evaluated with 30 listeners from University population who volunteered to participate in the experiment. The listeners included 18 males and 12 females between ages of 19 and 30. The average age of the participants was 24 (standard deviation 3.2). All participants reported normal hearing and no hearing difficulties.

The SRT was measured for two background noise conditions, at 50 kph and 100 kph, and defined as the SNR required for 50 % of sentences to be correctly identified, using the recommended procedure for the ‘hearing in noise test’ (HINT, [42]); scoring at the sentence level. Prior to the SRT evaluation, both speech and background noise recordings were needed to create a driving simulation, with speech and noise presented through headphones in the driving simulator. The binaural background noise recordings were obtained in a vehicle semi-anechoic four-wheel chassis dynamometer test chamber [17], both at a low speed (50 kph) and a high speed (100 kph) using a small sized sedan with a gasoline-fueled 1.8L inline 4-cylinder engine, and with a listener head and torso simulator (HATS 4100). The speed was adjusted using an actuator attached to the throttle pedal and the transmission selector was set to drive during the measurements to simulate actual vehicle operation. The background noise measurements did not account for the contribution of wind noise. Vehicle dynamometer rollers with textured coverings were utilized to simulate rough roadway surface conditions. The roller assembly noise rating was less than 50 dBA at 1 meter above the rolls at 60 miles per hour. The test chamber background noise rating was NC-20.

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 received speech inside an operating vehicle.

The speech material (HINT) sentences were prerecorded in the vehicle with the simulated talker (HATS) at three positions: passenger front (F), rear left (L), or rear right (R) seat. Prior to making the in-vehicle speech recordings, the HINT speech 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, and equalized in order to adjust its spectrum at a one-meter distance from the artificial mouth, using the IEC standard specification for male speech (IEC 60268-16: 2003, [36]). The spectrum levels of the HINT sentences delivered through the headphones to the human listeners operating a Brüel&Kjær NVH Desktop 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). These three configurations were used for both low- and high-speed conditions, for a total of six test conditions.

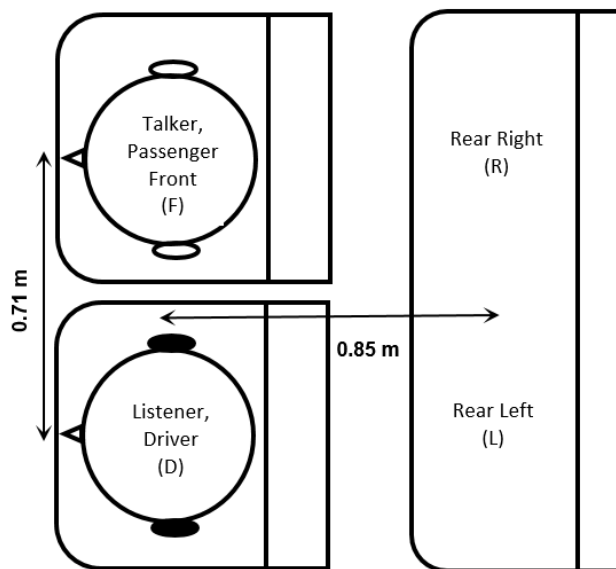


Figure 1: Vehicle speech and noise example measurement configuration with a talker tested in the passenger front position (F). Other possible talker locations were the passenger rear right (R) and rear left (L) positions. The listener was always simulated as the driver (D) position.

3.2. Objective metrics of in-vehicle speech intelligibility

According to the modeling procedure described in Sec. 2, to predict speech recognition performance, it is necessary to select an appropriate objective index of speech intelligibility to be implemented with the psychometric function. In the current analysis, two separate objective metrics were chosen based on previous modeling studies from our group (e.g., [15, 16]). These included the effective target-to-masker ratio (effective TMR) based on a model of binaural speech intelligibility proposed by Lavandier et al. (2012) [43] and the binaural loudness ratio (BLR) based on a model of binaural loudness [15].

Differences in effective TMR across tested conditions directly predict differences in SRT between these conditions. The model *lavandier2022* considers the benefit from binaural hearing in terms of the reduction in speech recognition threshold in dB relatively to a monaural or co-located condition [44]. This model is a slightly updated implementation of the model initially proposed by Lavandier & Culling (2010) [45] and it has been validated extensively in anechoic laboratory settings [46, 47], as well as in realistic reverberant conditions [43]. On the other hand, the BLR is the ratio of the binaural loudness of the speech to the binaural loudness of the noise, both specified in sones. Therefore, BLR is a dimensionless metric (sone/sone). The binaural loudness is estimated using the computational steps described by Moore et al. (2014) [48] and specified in ISO 532-2 (2017), [49]. These indices were selected because they involve modeling stages that are representative of binaural hearing, which is necessary to capture the effect of reverberation and spatial distribution of sound sources within a vehicle. Both the effective TMR and BLR have been previously applied to model the average speech recognition threshold for the in-vehicle test conditions considered here, with demonstrated success [16].

The effective TMR was computed following the procedure recommended for *lavandier2022* [44]. Briefly, the binaural recordings of the target speech and vehicle noise from the recognition experiment were inputs to the model. The model output two indices quantifying binaural hearing phenomena called binaural

unmasking and better-ear listening, the indices being combined to ultimately calculate TMR, also considering the speech intelligibility index (SII) weightings (ANSI S3.5, 1997, [35]).

The BLR was computed in two steps. First, the binaural loudness of the speech and noise were calculated separately using the procedure specified in ISO 532–2 (2017) [49], as implemented in the “Connect” software of Brüel&Kjær (Nærum, Denmark). Second, the ratio of these two quantities is taken, giving the BLR.

3.3 Analysis Procedures

To fit the psychometric function specified in Eq. 1 to the in-vehicle speech recognition data described in Sec. 3.1, the objective metrics (z_i) were first calculated for the stimuli used on each trial for each participant in the experiment. This was done separately for the effective TMR and BLR. In the original speech recognition test, the level of the background noise was calibrated to reflect the actual noise generated by the running vehicle in the test chamber. The level of the target speech was varied according to a staircase adaptive procedure from trial to trial, as recommended for the assessment of the SRT using the HINT sentences [42]. For each of the two objective metrics, all speech levels tested in all six conditions of the experiment were used to reconstruct the binaural speech stimuli for each participant. Along with the recorded binaural noise stimuli, the objective metrics were calculated for all trials, with one effective TMR and one BLR for each trial.

For each listener, one psychometric function was estimated for each of the two objective metrics based on all experimental trials across all six test conditions, excluding the first trial, considered to be a practice trial, for each condition. This led to a total of 120 trials of data contributing to the fitting of the psychometric function. For each of the effective TMR and BLR, one psychometric function was derived according to Eqs. 2-3 from the computed objective metrics (z_i) and the responses ($r_i = 0$ or 1) collected from the 120 trials using the Psignifit 4 Matlab Toolbox [50]. This was repeated for all 30 listeners.

3.4 Results

The fitted psychometric functions are shown in the top panel of Figure 2 for the two objective metrics. For the BLR (left), the psychometric function relates the BLR value to the proportion correct in sentence recognition. For the effective TMR (right), the psychometric function relates the effective TMR to the proportion correct in sentence recognition. Substantial individual differences are present in the psychometric functions. At the same BLR value or binaural effective TMR, speech recognition performance may vary widely among listeners. It seems that inter-subject variability in understanding speech in noise drives the variance in the current dataset. Individual listeners may not be treated as a random factor. The current analysis illustrates that individual listeners are often not homogeneous in their efficiency in performing a speech recognition task. The lower panels of Figure 2 show the histograms of the estimated threshold and slope parameters across the 30 listeners. For the BLR, the estimated threshold parameter has a mean of 0.67 and a standard deviation of 0.07, and the estimated slope parameter has a mean of 255.46%/unit BLR and a standard deviation of 87.79%/unit BLR. For the effective TMR, the estimated threshold parameter has a mean of 12.13 dB and a standard deviation of 1.53 dB, and the estimated slope parameter has a mean of 7.12 %/dB and a standard deviation of 1.91 %/dB.

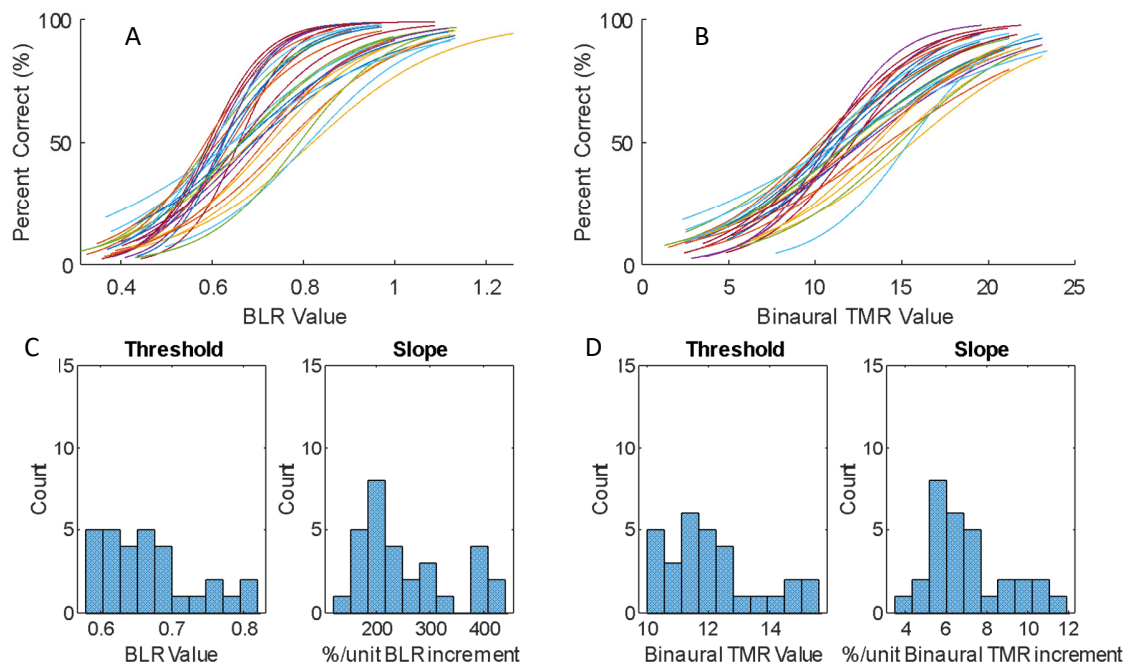


Figure 2: Panel A: the estimated individual psychometric functions that relate the BLR value to percentage correct. Each curve in the panel indicates the psychometric function for one listener, estimated by pooling data across the six test conditions. Panel B: Same as panel A, but for psychometric functions relate the binaural effective TMR to percentage correct. Panel C: the distributions of the estimated psychometric-function thresholds and slopes for BLR. Panel D: Same as panel C, but for binaural effective TMR.

Using all experimental trials averaged across all six test conditions, Figure 3 shows that the estimated individual BLR value and binaural effective TMR at a performance level of 50% are highly correlated [$r = 1.00, p < .001$]. Therefore, the threshold predictions from one model may be converted into predictions from the other using a simple linear function. The ability to quickly and conveniently calculate indices associated with a variety of validated speech intelligibility models is very useful for the in-vehicle speech intelligibility optimization.

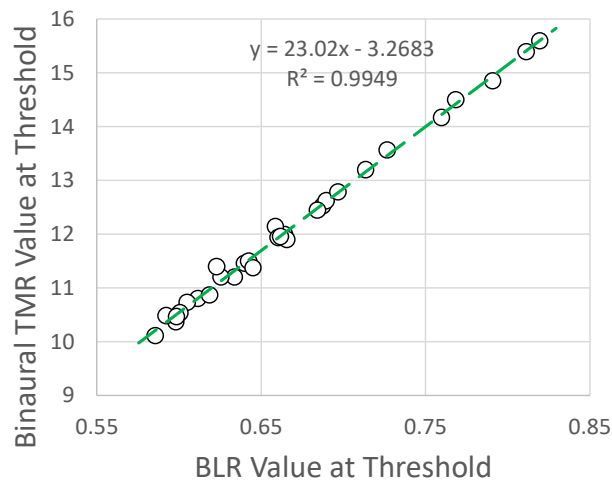


Figure 3: The estimated individual binaural effective TMR at the SRT for each of the 30 listeners plotted against the estimated individual BLR values at threshold.

4. Discussion and Conclusion

In the current study, a modeling technique is presented to fit psychometric functions to in-vehicle speech recognition data. This approach combines an objective metric of speech intelligibility in the acoustic environment of interest and a psychometric function that links the objective metric to the speech recognition accuracy. The acoustic factors are being modeled by the objective metric based on the in-situ recordings of the target speech and background noise, while the factors inherent to individual listeners are being modeled by the threshold and slope parameters of the psychometric function. This modeling technique can be applied to any speech-recognition data where the trial-by-trial speech and noise stimuli and the listener's responses are available. For modeling in-vehicle speech intelligibility, it is typically possible to reconstruct the speech and noise stimuli based on the impulse responses measured from the source to the listener locations. This is illustrated in the current study using previously collected in-vehicle speech recognition data [17].

Through this demonstration, it is shown that the fitted psychometric functions can link an objective metric of speech intelligibility to the speech recognition performance in terms of proportion correct. This means that once fitted, the fitted model is capable of predicting the listener's speech recognition performance for a novel unseen acoustic situation. For example, it is possible to predict the anticipated benefit of a sound treatment applied to a vehicle in terms of an improvement in proportion correct. Similarly, the fitted model can be used to determine the suitable acoustic designs that would meet a certain minimum speech recognition accuracy for the listener. Of course, this approach assumes that the binaural front-end model is able to accurately predict the influence of the acoustic conditions being considered. If this assumption cannot be made based on previous model validation, then this front-end model associated with the psychometric function will probably not fit the data well, and there would be no way to tell whether it is caused by the front-end model or the psychometric curve. The estimated threshold and slope parameter may not reflect the characteristics of the listeners but be dominated by errors raised from poor model fits. Therefore, only objective indexes that have been validated for comparable test conditions are suitable for the current analysis approach. For example, for the vehicle dynamometer condition used in this study, both BLR and LAV models resulted in comparable SRT predictions, however, the BLR model has been shown

to fail to predict SRTs when binaural effects are involved [16], so that it should not be used then. The outcome that binaural effects adversely affect BLR predictions was expected considering a review and comparison of existing binaural speech intelligibility models, including their domain of validity, as described by Lavandier and Best, 2020 [51]. This review highlights that it is crucial to consider better-ear listening and binaural unmasking for accurate binaural intelligibility predictions, and BLR does not model these effects.

When multiple speech-recognition models are compared, coupling these models with psychometric functions enabled direct predictions of speech-recognition scores regardless of the specific output variables or units of the models (i.e. sone/sone in the BLR model, and dB for TMR in the LAV model). This means that (1) each of the models can be used to fit the trial-by-trial data, (2) the fitted models can be directly compared in terms of their likelihood, (3) the model predictions can be compared at performance levels other than 50%. Importantly, fitting psychometric functions to the same data set but with different front-end models makes sense only when both models accurately predict the same data; only then it is possible to calibrate/translate the outputs of the front-end models, usually expressed in different units.

More importantly, the current study highlights the large individual differences among listeners, exposing it as a potential issue when optimizing vehicle interior speech intelligibility based on a “model/average listener”. The cause of the individual differences is complex and related to variations in peripheral auditory, central auditory, and cognitive processes [52 – 57]. Additionally, by conducting behavioral evaluations of in-vehicle speech recognition, as for the data set used here, the current modeling technique allows the establishment of the distribution of the psychometric-function parameters. From this, the range of anticipated speech-recognition performances can be predicted. Based on the derived distribution, it is possible to determine the proportion of listeners that can achieve a certain level of speech-recognition performance for a given acoustic environment.

Despite the demonstrated promise of the presented modeling approach, there are a number of limitations that may be worth considering. First, the threshold and slope parameters of the psychometric

function depend not only on the inherent speech-understanding efficiency of the listener, but also on the material used [27]. Therefore, to enable direct comparisons of the psychometric functions across different work sites and vehicle manufacturers, it is necessary to use standardized testing materials for the behavioral experiment. Whether the HINT sentences are the most suitable for this purpose is not clear, and the rationale for selecting an appropriate speech material is beyond the scope of the current study. Second, in the current formulation of the psychometric function (Eq. 1), the threshold and slope parameters are allowed to vary from one listener to another, but assumed independent of test conditions (the influence of which is assumed to be properly captured by a binaural front-end model, previously validated for such conditions). Therefore, the current model implementations separated the effects of test condition and individual listeners into the front-end objective metric and the psychometric-function parameters, respectively. However, this approach omits the facts that listeners also vary in their sensory processing of the acoustic inputs. This can be in part considered by the model (before the psychometric function). For example, Vicente et al. (2020) [58] adapted the model *lavandier2022* to take into account the effect of hearing loss and make intelligibility predictions for hearing-impaired listeners, thus modelling part of the individual variability. Lavandier et al. (2021) [19] showed that this does not consider all the individual variability, in particular for normal-hearing listeners. In future updates, it is possible to fit additional parameters of the objective metric that represent peripheral and binaural processing to individual listeners' data. This would help capture the potential individual differences not only in speech understanding efficiency but also in auditory capabilities.

References

- [1] Li, Y., Monroe, R., and Geist, B. (2023b). Constant Power Psychoacoustic Spectrum Optimization for Loudness and Sharpness with Application to Vehicle Interiors. *SAE Int. J. Veh. Dyn., Stab., and NVH* 7(2):173-193.
- [2] Li, T., Li, P., Jiang, D., and Qiu, B. (2023) Squeak and Rattle Analysis of Automotive Interiors Considering Contact Nonlinearity. *SAE Int. J. Veh. Dyn., Stab., and NVH* 7(2):209-220.
- [3] Chisolm, T.H., Noe, C.M., McArdle, R., Abrams, H. (2007). Evidence for the use of hearing assistive technology by adults: The role of the FM system' *Trends in Amplification*. 11(2), 73-89.
- [4] Gatehouse, S., Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *Internat. J. Audiol*, 43(2), 85-99.
- [5] Johnson J.E. (1995). Rural elders and the decision to stop driving. *J Comm. Health Nurs.*, 12(3), 131-138.
- [6] Kochkin S (2005). MarkeTrak VII: Hearing loss population tops 31 million. *Hear Rev*, 12(7), 16-29.
- [7] Granat, C. (2008). Using the Speech Transmission Index for a Better Analysis of the Speech Transmission Inside Vehicle. *NOISE-CON 2008*, Dearborn, Michigan.
- [8] Samardzic, N. and Novak, C. (2011). In-vehicle speech intelligibility for different driving conditions using the speech transmission index. *Noise Control Engineering Journal*, Vol. 59, No. 4, pp. 397-407.
- [9] Samardzic, N. and Novak, C. (2011b). In-vehicle application of common speech intelligibility metrics', *Int. J. Vehicle Noise and Vibration*. Vol. 7, No. 4, pp. 328-346.
- [10] 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 based on On-Road Measurements and the Associated Contributions from the Ordered and Masking Noise Sources. *J. Audio Eng. Soc* Vol. 62, No. 9, pp. 676-687.
- [11] Samardzic, N. (2014) 'The Applicability of the Objective Speech Intelligibility Metrics for Vehicle Interior Speech Intelligibility Evaluation, Considering Different Listening Configurations and Background Noise Spectra', *SAE Int. J. Passeng. Cars - Mech. Syst.* Vol. 7, No. 1, pp. 434-438.
- [12] Ebbitt, G. L. and Remtema, T. M. (2017). Automotive speech intelligibility measurements. *Sound & Vib.*, 51(6), 6–11.
- [13] Liang, L., & Yu, G. (2020). Binaural speech transmission index with spatialized virtual speaker in near field: Distance and direction dependence. *The Journal of the Acoustical Society of America*, 148(2), EL202.
- [14] Rouch, J., & Parizet, E. (2018). Speech Modifications to Increase the Intelligibility of Vocal Messages Broadcast by Driving Assistance Systems Intended For Hearing-Impaired Drivers. *Acta Acustica united with Acustica*, 104(4), 668-677.
- [15] Samardzic, N. & Moore, B.C.J (2021). Binaural speech-to-noise ratio at the speech recognition threshold in vehicles. *Noise Control Engineering Journal*, 69(2), 173–178.
- [16] Samardzic, N. & Lavandier, M. (2023). Further evaluation of the binaural loudness ratio to predict speech intelligibility in vehicles in rooms. *Applied Acoustics*, 203, 109140.

- [17] Samardzic N, Novak C, 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*, 8:318–36.
- [18] Schoof, T., & Rosen, S. (2014). The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. *Frontiers in aging neuroscience*, 6, 307.
- [19] Lavandier, M., Mason, C. R., Baltzell L. S., Best, V. (2021) Individual differences in speech intelligibility at a cocktail party: A modelling perspective. *J. Acoust. Soc. Am.*, 150 (2), 1076-1087.
- [20] Shen, Y. & Richards, V. (2012). A maximum-likelihood procedure for estimating psychometric functions: Thresholds, slopes, and lapses of attention, *J Acoust Soc Am*, 132(2): 957–967.
- [21] Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & psychophysics*, 63(8), 1421-1455.
- [22] Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8), 1293-1313.
- [23] Shen, Y., Manzano, N. K., & Richards, V. M. (2015). Psychometric functions for sentence recognition in sinusoidally amplitude-modulated noises. *J Acoust Soc Am*, 138(6), 3613-3624.
- [24] Ross, B., Dobri, S., & Schumann, A. (2021). Psychometric function for speech-in-noise tests accounts for word-recognition deficits in older listeners. *J Acoust Soc Am*, 149(4), 2337-2352.
- [25] Wilson, R. H., & Oyler, A. L. (1997). Psychometric functions for the CID W-22 and NU Auditory Test No. 6. Materials spoken by the same speaker. *Ear and Hearing*, 18(5), 430-433.
- [26] Young Jr, L. L., Dudley, B., & Gunter, M. B. (1982). Thresholds and psychometric functions of the individual spondaic words. *Journal of Speech, Language, and Hearing Research*, 25(4), 586-593.
- [27] MacPherson, A., & Akeroyd, M. A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey. *Trends in Hearing*, 18, 2331216514537722.
- [28] Relaño-Iborra, H., May, T., Zaar, J., Scheidiger, C., & Dau, T. (2016). Predicting speech intelligibility based on a correlation metric in the envelope power spectrum domain. *The Journal of the Acoustical Society of America*, 140(4), 2670-2679.
- [29] Hu, W., Swanson, B. A., & Heller, G. Z. (2015). A statistical method for the analysis of speech intelligibility tests. *PloS one*, 10(7), e0132409.
- [30] Rhebergen, K. S., Maalderink, T. H., & Dreschler, W. A. (2017). Characterizing speech intelligibility in noise after wide dynamic range compression. *Ear and Hearing*, 38(2), 194-204.
- [31] Kates, J. M., & Arehart, K. H. (2022). An overview of the HASPI and HASQI metrics for predicting speech intelligibility and speech quality for normal hearing, hearing loss, and hearing aids. *Hearing Research*, 108608.
- [32] Spille, C., Ewert, S. D., Kollmeier, B., & Meyer, B. T. (2018). Predicting speech intelligibility with deep neural networks. *Computer Speech & Language*, 48, 51-66.
- [33] Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., & Francart, T. (2018). Speech intelligibility predicted from neural entrainment of the speech envelope. *Journal of the Association for Research in Otolaryngology*, 19, 181-191.

- [34] Sobon, K. A., Taleb, N. M., Buss, E., Grose, J. H., & Calandruccio, L. (2019). Psychometric function slope for speech-in-noise and speech-in-speech: Effects of development and aging. *The Journal of the Acoustical Society of America*, 145(4), EL284-EL290.
- [35] ANSI. ANSI S3.5-1997 (2017). American National Standard Methods for the Calculation of the Speech Intelligibility Index (American National Standards Institute, New York).
- [36] IEC 60268-16: 2003 (2003). Sound System Equipment - Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index, International Electrotechnical Commission, Geneva, Switzerland.
- [37] Taal, Cees & Hendriks, Richard & Heusdens, R. & Jensen, Jesper. (2011). An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech. *Audio, Speech, and Language Processing*, IEEE Transactions on. 19. 2125 - 2136. 10.1109/TASL.2011.2114881.
- [38] Treutwein, B., & Strasburger, H. (1999). Fitting the psychometric function. *Perception & psychophysics*, 61(1), 87-106.
- [39] Kuss, M., Jäkel, F., & Wichmann, F. A. (2005). Bayesian inference for psychometric functions. *Journal of Vision*, 5(5), 8-8.
- [40] Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America*, 111(6), 2801-2810.
- [41] Allen, P., & Wightman, F. (1994). Psychometric functions for children's detection of tones in noise. *Journal of Speech, Language, and Hearing Research*, 37(1), 205-215.
- [42] Nilsson, M., Soli, S.D., & Sullivan, J.A. (1994). Development of the Hearing in Noise Test for the measurement of speech recognition thresholds in quiet and in noise. *J. Acoust. Soc. Am.*, 95, 1085– 1099.
- [43] Lavandier M., Jelfs S., Culling J.F., Watkins A.J., Raimond A.P., Makin S.J. (2012) Binaural prediction of speech intelligibility in reverberant rooms with multiple noise sources. *J Acoust Soc Am*, 131:218–31.
- [44] Lavandier, M., Vicente, T. & Prud'homme L. (2022) A series of SNR-based speech intelligibility models in the Auditory Modeling Toolbox. *Acta Acustica*, 6 (20):1-17.
- [45] Lavandier, M. & Culling JF. (2010). Prediction of binaural speech intelligibility against noise in rooms. *J Acoust Soc Am*, 127(1):387–99.
- [46] Culling J.F., Lavandier M. & Jelfs S. (2013). Predicting binaural speech intelligibility in architectural acoustics. In: Blauert J, editor. *The Technology of Binaural Listening*. Berlin-Heidelberg-New York, NY: Springer; 427–47.
- [47] Jelfs, S., Culling, J. F., & Lavandier, M. (2011). Revision and validation of a binaural model for speech intelligibility in noise. *Hearing research*, 275(1-2), 96–104.
- [48] Moore B. C. (2014). Development and current status of the "Cambridge" loudness models. *Trends in Hearing*, 18.
- [49] ISO 532-2:2017 (2017) Acoustics — Methods for calculating loudness - Part 2: Moore-Glasberg method, International Standard, International Organization for Standardization, Geneva, Switzerland.
- [50] Schütt, H., Harmeling, S., Macke, J., & Wichmann, F. (2015). Psignifit 4: Pain-free Bayesian inference for psychometric functions. *Journal of vision*, 15(12), 474-474.
- [51] Lavandier M. and Best V. (2020). Modeling binaural speech understanding in complex situations. In: *The Technology of Binaural Understanding*, Blauert J and Braasch J (eds.), Switzerland: Springer, Chapter 19, 547-578.

- [52] Humes, L. E., Kidd, G. R., & Lentz, J. J. (2013). Auditory and cognitive factors underlying individual differences in aided speech-understanding among older adults. *Frontiers in systems neuroscience*, 7, 55.
- [53] Füllgrabe, C., Moore, B. C., & Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Frontiers in aging neuroscience*, 6, 347.
- [54] Schoof, T., & Rosen, S. (2014). The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. *Frontiers in aging neuroscience*, 6, 307.
- [55] Humes, L. E., & Dubno, J. R. (2010). Factors affecting speech understanding in older adults. *The aging auditory system*, 211-257.
- [56] Gordon-Salant, S., & Fitzgibbons, P. J. (1997). Selected cognitive factors and speech recognition performance among young and elderly listeners. *Journal of Speech, Language, and Hearing Research*, 40(2), 423-431.
- [57] George, E. L., Festen, J. M., & Houtgast, T. (2006). Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical society of America*, 120(4), 2295-2311.
- [58] Vicente, T., Lavandier, M., Buchholz, J. M. (2020). A binaural model implementing an internal noise to predict the effect of hearing impairment on speech intelligibility in non-stationary noises. *J. Acoust. Soc. Am.*, 148 (5), 3305-3317.