Binaural speech-to-noise loudness ratio at the speech reception threshold in vehicles

Nikolina Samardzic [a] and Brian C.J. Moore [b]

(Received: 17 March 2020; Revised:; Accepted:)

Traditional methods for predicting the intelligibility of speech in the presence of noise inside a vehicle, such as the Articulation Index (AI), the Speech Intelligibility Index (SII), and the Speech Transmission Index (STI), are not accurate, probably because they do not take binaural listening into account; the signals reaching the two ears can differ markedly depending on the positions of the talker and listener. We propose a new method for predicting the intelligibility of speech in a vehicle, based on the ratio of the binaural loudness of the speech to the binaural loudness of the noise, each calculated using the method specified in ISO 532-2 (2017). The method was found to give accurate predictions of the Speech Reception Threshold (SRT) measured under a variety of conditions and for different positions of the talker and listener in a car. The typical error in the predicted SRT was 1.3 dB, which is markedly smaller than estimated using the SII and STI (2.0 dB and 2.1 dB, respectively).

Primary subject classification: 63.3; Secondary subject classification: 13.2.1

1 INTRODUCTION

Several methods have been proposed for predicting the intelligibility of speech in quiet and in the presence of steady background sounds, based on measurements of the physical characteristics of the speech and background. These include the Articulation Index (AI) [1,2,3], the Speech Intelligibility Index (SII) [4,5] and the Speech Transmission Index (STI) [6,7]. While these methods are often reasonably accurate when the signals at the two ears are the same, it is difficult to know how to apply them when the signals are different at the two ears. One situation that has proved to be problematic is listening to speech in a moving car, which can be challenging for both normal-hearing and hearing-impaired listeners [8].

The methods mentioned above are all based on the long-term characteristics of the speech and background, and do not take into account the short-term properties of the sounds or differences in the sounds arriving at the two ears. Several more recent models have been proposed that take one or both of these factors into account. Rhebergen and Versfeld [9] developed a method based on the short-term SII. The speech and background are analyzed in small time frames to calculate the momentary SII and predict the amount of speech information available to the listener for that time frame. The SII values are then averaged across frames. The model gave reasonably accurate predictions of speech reception thresholds (SRTs, defined as the speech-to-background ratio needed to achieve a certain level of intelligibility, such as 50% correct) for speech in stationary noise [9], fluctuating noise [9,10], interrupted noise [11], and multiple-

a) Department of Engineering Technology, Lawrence Technological University, Southfield, MI 48075, USA; email: nsamardzi@ltu.edu.
b) Department of Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK; email: bcjm@cam.ac.uk.

talker noise [11]. However, the model does not accommodate situations where the speech and background differ across the two ears. Several other models for predicting the intelligibility of speech in fluctuating backgrounds also do not accommodate such situations [12,13,14].

Some models do accommodate situations where the speech and background differ across the two ears. The binaural speech intelligibility model of Beutelman and Brand [15] and Beutelmann *et al.* [16] is based on a combination of the equalization-cancellation model of Durlach [17] and the SII metric and it takes the short-term properties of the speech and background into account. It gave reasonably accurate predictions of SRTs for speech in a variety of simulated environments, including an office and a cafeteria. Lavandier and Culling [18] described a model in which binaural effects were modeled by computing binaural masking level differences [17] and monaural effects were predicted from the excitation patterns of the speech and noise. The model gave reasonably accurate predictions of SRTs for speech in steady background noise but it does not take into account the effects of short-term fluctuations in the sounds, so it is not accurate for fluctuating background sounds, such as speech. The models incorporating binaural effects are computationally intensive and require measurements of the time waveforms of the at-ear signals, separately for the speech and the background sounds. They have not, to our knowledge, been evaluated in the context of listening in a vehicle.

In this paper we focus on simpler and less computationally intensive models for predicting the intelligibility of speech in a vehicle, based on the long-term average characteristics of the speech and the background. This seems appropriate since, for a fixed vehicle and wind speed, the noise inside a moving vehicle is essentially steady. The paper starts by reviewing the ability of existing standardized metrics to predict the intelligibility of speech in a moving car and then describes a new metric, based on the binaural loudness ratio of the speech and the background sound, calculated using the loudness model described in ISO 532-2 [19]. This model was developed from the model of Moore *et al.* [20], but modified as described by Moore and Glasberg [21] to accommodate situations where the signals differ at the two ears. It is shown that this new model gives accurate predictions of the SRT for speech in a moving car under a variety of conditions.

The present analysis utilizes the raw data, specifically the speech and noise spectra at the SRT, of Samardzic *et al.* [22]. In the present study, the SII and STI were calculated using the spectrum of the speech material actually used for the SRT measurements, whereas previously [23,24,25] the SII metric was calculated based on the speech spectrum for normal vocal effort, as specified in the ANSI standard for the SII [5]. The calculations of the STI at the SRT (Table 2) have not previously been published.

2 PREDICTIONS OF TRADITIONAL SPEECH INTELLIGIBILITY METRICS

2.1 Predicted metric values for each ear for a listener in a car

Several traditional methods for predicting speech intelligibility require specification of the speech spectrum, the noise spectrum, and the hearing threshold levels of the listener, if the listener is hearing impaired. The most widely used speech intelligibility metric, the SII [5], utilizes these measurements. There are several variants of the method, but all involve calculation of the audibility of the speech in each of several frequency bands, and summation of audibility contributions across bands, with a weighting for each band according to its relative importance for speech intelligibility. The value of the SII ranges from 0, indicating that the speech is almost completely unintelligible, to 1, indicating excellent intelligibility. Other objective speech intelligibility metrics providing scores ranging from 0 to 1 are the AI [1,2,3] and the

STI [6,7]. In the automotive industry, a simplified and modified version of the AI method has often been used [26,27,28], referred to as the "vehicle AI" and sometimes attributed to Beranek [29]. This method neglects the effects of absolute threshold, since in a vehicle the sound levels are typically well above the absolute threshold over the frequency range that is important for speech intelligibility. A difference in vehicle AI scores of 0.06 (6%) or more is considered to be significant [30].

The STI is more accurate than the AI or SII in reverberant environments because it takes into account the effects of both noise and reverberation on speech intelligibility [6]. However, the STI is also more measurement intensive, as it involves measurements of amplitude modulation transfer functions [6] or impulse responses [31,32] to characterize reverberation. For listening in vehicles, the effects of reverberation are negligible and speech intelligibility, as quantified by the STI, is primarily limited by the long-term average characteristics of the background noise [23].

For the SII, scores > 0.75 are usually taken as indicating good speech communication, while poor communication systems have an SII < 0.45 [5]. STI scores for speech in car noise have been interpreted in the following way: > 0.75 - good or excellent speech communication; 0.6 to 0.75 - good communication; 0.45 to 0.6 – fair communication; 0.3 to 0.45 - poor communication; <0.3 - almost no intelligibility [23,24,25].

Figure 1 shows the experimental setups, using a mid-sized sedan, for speech signal calibration and semi-anechoic vehicle dynamometer speech and noise measurements used in a study of Samardzic and Novak [24]. A head and torso simulator (HATS) was used to estimate the spectra of the speech and noise at the listener's ears. The metrics were calculated based on the measured long-term average spectrum of the "hearing in noise test" (HINT) speech material [33] that was used in those studies.

Table 1 shows values of the SII and STI for each ear of the listener for several talker locations in a simulated car environment (dynamometer). The listener was always located in the front-left (driver's) seat. It is clear that for a fixed listening situation the values of the metrics can differ considerably across ears. For example, for a simulated vehicle speed of 100 km/hr and a talker located in the front-right seat, the SII was 0.4 for the left ear and 0.6 for the right ear, while the STI was 0.43 for the left ear and 0.66 for the right ear. The SII and the STI do not take into account the effects of differences in the speech and noise at the two ears. These differences are often considerable [34]. Also, there are no recommendations about how to combine metric values across ears. This may be one factor that contributes to the finding that these metrics do not give accurate predictions of the intelligibility of speech in vehicle noise [23,24,25,35,36,37].

Genuit [38] argued that speech communication in a noisy environment depends strongly on binaural processing. This processing supports directional hearing and binaural release from masking[39,40]. In addition, the outer ear acts as a directional filter that can change the sound pressure level at the eardrum by +15 to −30 dB, depending on the frequency and direction of sound incidence [34], and this can introduce strong inter-aural level differences. To accurately characterize the signals reaching each ear, it is essential to perform binaural measurements either using in-ear microphones or, as in the study of Samardzic and Novak [24], using a HATS.

2.2 Relationship between objective and subjective measures of speech intelligibility

Samardzic *et al.* [25] assessed the intelligibility of speech in the presence of vehicle noise for various conditions using the HINT [33]. They measured the SRT for each condition, defined as the speech-to-noise ratio (SNR) required for 50% of sentences to be correctly identified, using the recommended procedure for the HINT and 35 normal-hearing subjects. The HINT sentences were prerecorded in the vehicle with the simulated talker (HATS) at a variety of positions. Similarly, the background sounds, produced both on-road and in the vehicle dynamometer test chamber, were recorded (separately) using a HATS for various positions and various conditions. The recordings were later presented via headphones in the driving simulator, as illustrated in Fig. 2. 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 levels in the vehicle. The spectrum levels of the HINT sentences delivered through the headphones to the human subjects in the driving simulator matched those recorded inside the vehicle to within 0.2 dB, as verified using the HATS, and Bruel and Kjaer NVH Simulator software. The driving simulator is illustrated in Fig. 2.

In theory, if a given intelligibility metric is accurate, the value of that metric at the SRT should be constant across conditions. However, as noted earlier, the traditional metrics such as the AI, SII, and STI give a separate value for each ear, and it is not obvious how metric values should be combined across ears. In section 3, metric values for the SII and STI are presented for the ear that was closer to the talker, based on the idea that when the speech-to-background ratio differs across ears, intelligibility is dominated by the ear receiving the higher speech-to-background ratio [39].

3 THE BINAURAL LOUDNESS RATIO AT THE SRT

In this section a new speech intelligibility metric is introduced and compared to the SII and STI metrics, using the same dataset of twelve in-vehicle configurations of talker and listener.

3.1 Method

The speech intelligibility metric proposed here 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. We refer to this ratio as the Binaural Loudness Ratio (BLR). The loudness of the speech and noise were calculated separately using the procedure specified in ISO 532-2 [19], as implemented in the "Connect" software of Brüel & Kjær (Nærum, Denmark). This procedure is based on the model described by Moore *et al.* [20], but modified to take into account the finding that loudness does not simply sum across ears; rather, a diotic sound is about 1.5 times as loud as that same sound presented monaurally [21,42]. Moore and Glasberg [21] modeled this finding using the concept of binaural inhibition, namely that the internal representation of the signal at one ear can be reduced (inhibited) by a signal at the other ear [43]. Briefly, the 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 [40,44]; (3) Transformation of the frequency scale to the ERB_N-number scale, which approximates the way that frequency is mapped to place within the cochlea [44]; (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; (6) Summation across center frequencies and across ears to give the overall predicted loudness. For more details see Moore [41].

The proposed method has the following steps: (1) The binaural loudness of the speech is calculated from the long-term average of the 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; (3) The ratio of these two quantities is taken, giving the BLR. The method was applied to stimuli for the same configurations of talker and listener locations and background noise conditions as previously used for the measurement of SRTs [25] and listed in Table 2.

3.2 Results

The outcomes are shown in Table 2. The average BLR across all conditions, was 0.643, with a standard deviation (SD) of 0.059. The SD was about 9% of the mean. In other words, the calculated BLR was roughly constant at the SRT, consistent with the idea that the SRT corresponds roughly to a constant ratio of the loudness of the speech and the noise. To assess the importance of the deviations of the BLR from 0.643 for the individual conditions, we calculated the amount by which the SNR needed to be adjusted in order for the BLR to reach the average value of 0.643. These adjustments indicate the effective error that would occur if the SRT were predicted from the BLR. The magnitudes of the adjustments for each condition are shown in Table 2. The adjustments range from −2.6 to 1.8 dB. The SD of the adjustments was 1.3 dB. Thus, the SRT could be predicted from the BLR with a typical error of only 1.3 dB.

Table 2 also shows the outcomes for the SII and STI. The average SII value across all conditions, was 0.394, with an SD of 0.066. The SD was about 17% of the mean. We calculated the amount by which the SNR needed to be adjusted in order for the SII to reach the average value of 0.394. The magnitudes of the adjustments for each condition are shown in Table II. The adjustments ranged from −3.2 to 2.9 dB. The SD of the adjustments was 2.0 dB, which is larger than for the BLR. We conclude that the SII for the ear closer to the talker gives less accurate predictions of the SRT than the BLR. The average STI value across all conditions, was 0.390, with an SD of 0.069. The SD was about 18% of the mean. We calculated the amount by which the SNR needed to be adjusted in order for the STI to reach the average value of 0.390. The adjustments ranged from −3.4 to 3.0 dB. The SD of the adjustments was 2.1 dB, which is larger than for the BLR. We conclude that the STI for the ear closer to the talker also gives less accurate predictions of the SRT than the BLR.

4 DISCUSSION

The SII, the STI, and the BLR are calculated from the long-term average properties of the speech and the background noise. Thus fluctuations in the speech and background are ignored. This approach makes all metrics quick and easy to calculate. Despite the simplicity of the BLR metric, it performed surprisingly well in predicting the SRTs under a variety of conditions, the SD of the errors being only 1.3 dB. In contrast, the SII and STI led to less accurate predictions, the SDs of the errors being 2.0 and 2.1 dB, respectively. The poorer accuracy of the SII and STI probably stems at least partly from the fact that binaural processing is not taken into account and there is no standard way to combine the metrics' values at the two ears when those values differ, as is often the case when listening in a vehicle.

An important aspect of the loudness model on which the BLR calculations were based is that it allows the exact specification of the spectra of the signals reaching the two ears of the listener, and hence it includes effects such as the directivity and orientation of the talker, the transmission of sound from the talker to the listener, and the acoustic effects of the torso, head, and pinnae. This is important, as these effects can significantly influence the spectra of the signal and background at the two ears. The HATS used to estimate the at-ear signals is intended to have acoustical properties representative of a typical human head, and given that 35 listeners were tested, it seems reasonable to assume that it gave an accurate representation of the average spectra at the listeners' ears. However, individual differences in head-related transfer functions might lead to individual differences in the measured and predicted SRTs. This is a topic for future study.

It should be noted that the SRTs used to evaluate the BLR as a predictor of speech intelligibility were obtained in simulated driving conditions, where both visual and haptic feedback were provided. The effects of using such realistic conditions are significant; the same subjects required on average about a 3-dB higher SNR to reach the SRT while driving and listening than when just listening [25]. Additionally, there was an increase in the SD of the SRTs across subjects of 0.5 dB with the inclusion of the visual and haptic feedback. It would be of interest in future studies to evaluate the BLR as a predictor of speech intelligibility under conditions of "pure" listening, without simulated driving. This would provide a solid foundation for subsequent systematic quantification of additional effects associated with multisensory processing in simulated or real complex environments.

It remains unknown whether the BLR gives accurate predictions of SRTs for steady background sounds other than vehicle noise. It also remains unknown whether the accuracy of the predictions could be improved using a loudness model for time-varying sounds [45,46] or using a model for predicting the partial loudness of time-varying sounds in time-varying backgrounds [47]. These are topics for future research.

5 CONCLUSIONS

A new method has been proposed for predicting the intelligibility of speech in the steady background noise of a car under conditions of binaural listening, using the long-term average spectra of the speech and the background noise at each ear of the listener. The method uses as a predictor the binaural loudness ratio (BLR) of the speech alone and the noise alone, calculated using a loudness model (ISO 532-2, 2017). The method predicted the SRT for a variety of talker/listener configurations and different car speeds with good accuracy, the SD of the errors being only 1.3 dB.

6 ACKNOWLEDGMENTS

We thank three reviewers for helpful comments on an earlier version of this paper.

7 REFERENCES

1. N. R. French and J. C., Steinberg. "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.*, **19**, 90-119, (1947).
2. H. Fletcher. *Speech and Hearing in Communication*, Van Nostrand, New York, 1-461, (1953).
3. *Methods for the calculation of the articulation index*, American National Standards Institute ANSI S3.5-1969, New York, (1969).
4. C. Pavlovic. "Derivation of primary parameters and procedures for use in speech intelligibility predictions," *J. Acoust. Soc. Am.*, **82**, 413-422, (1987).

5. *Methods for the calculation of the speech intelligibility index*, American National Standards Institute ANSI S3.5-1997, New York (2017).
6. H. J. M. Steeneken and T. Houtgast. "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Am.*, **67**, 318-326 (1980).
7. *Sound System Equipment - Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index*, International Electrotechnical Commission, IEC 60268-16: 2003, Geneva, Switzerland (2003).
8. V. M. Richards, B. C. J. Moore and S. Launer. "Potential benefits of across-aid communication for bilaterally aided people: Listening in a car," *Int. J. Audiol.*, **45**, 182-189 (2006).
9. K.S. Rhebergen and N.J. Versfeld. "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**, 2181-2192 (2005).
10. K.S. Rhebergen, N.J. Versfeld, and W.A. Dreschler. "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," *J. Acoust. Soc. Am.*, **120**, 3988-3997 (2006).
11. K.S. Rhebergen, N.J. Versfeld, W.A. Dreschler. "The dynamic range of speech, compression, and its effect on the speech reception threshold in stationary and interrupted noise," *J. Acoust. Soc. Am.*, **126**, 3236-3245 (2009).
12. S. Jørgensen, S.D. Ewert, and T. Dau. "A multi-resolution envelope-power based model for speech intelligibility," *J. Acoust. Soc. Am.*, **134**, 436-446 (2013).
13. H. Relano-Iborra, T. May, J. Zaar, C. Scheidiger, and T. Dau. "Predicting speech intelligibility based on a correlation metric in the envelope power spectrum domain," *J. Acoust. Soc. Am.*, **140**, 2670-2679 (2016).
14. T. Biberger and S.D. Ewert. "The role of short-time intensity and envelope power for speech intelligibility and psychoacoustic masking," *J. Acoust. Soc. Am.*, **142**, 1098-1111 (2017).
15. R. Beutelmann and T. Brand. "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.*, **120**, 331-342 (2006).
16. R. Beutelmann, T. Brand, and B. Kollmeier. "Revision, extension, and evaluation of a binaural speech intelligibility model," *J. Acoust. Soc.Am.*, **127**, 2479-2497 (2010).
17. N.I. Durlach. “Equalization and cancellation theory of binaural masking-level differences,” *J. Acoust. Soc. Am.*, **35**, 1206–1218 (1963).
18. M. Lavandier and J.F. Culling. “Prediction of binaural speech intelligibility against noise in rooms” *J. Acoust. Soc. Am.*, **127**, 387-399 (2010).
19. *Acoustics - Methods for calculating loudness - Part 2: Moore-Glasberg method*, International Standard ISO 532-2:2017, International Organization for Standardization, Geneva, Switzerland (2017).
20. B.C.J. Moore, B.R. Glasberg, and T. Baer. "A model for the prediction of thresholds, loudness and partial loudness," *J. Audio Eng. Soc.*, **45**, 224-240 (1997).
21. B.C.J. Moore and B.R. Glasberg. "Modeling binaural loudness," *J. Acoust. Soc. Am.*, **121**, 1604-1612 (2007).

22. C. Granat. "Using the Speech Transmission Index for a better analysis of the speech transmission inside vehicle," in *Proc. NOISE-CON 2008*, Dearborn, MI (2008).
23. N. Samardzic and C. Novak. "In-vehicle speech intelligibility for different driving conditions using the Speech Transmission Index," *Noise Control Eng. J.*, **59**, 397-407 (2011).
24. N. Samardzic and C. Novak. "In-vehicle application of common speech intelligibility metrics," *Int. J. Vehicle Noise. Vib.*, **7**, 328-346 (2011).
25. N. Samardzic, C. Novak, and R. Gaspar. "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 (2012).
26. N. Hilyard and A. Cunningham. "Design optimisation of MDI-based polyurethane foam backed automotive carpet systems," *SAE Technical Paper 911087*, SAE, Warrendale, PA (1991).
27. T. Connelly, J. Knittel, and M. Jay. "A study of NVH vehicle testing variability," *SAE J. . Pass. Cars; Mech. Sys.*, **114**, 2838-2847 (2005).
28. J. Lu. "Noise reduction and sound quality improvement with acoustic windshield," in *Proc. SAE Brasil Noise and Vibration Conference*, SAE, 2008-36-0507 (2008).
29. L.L. Beranek. "The design of speech communication systems," in *Proc. Inst. Rad. Eng.*, **35**, 880-890 (1947).
30. G. Ebbitt. "The assessment of speech intelligibility in vehicle " in *Proc. NOISE-CON 2001*, Portland, Maine (2001).
31. F. Bozzoli, and A. Farina. "Measurement of Speech Transmission Index inside cars using throat-activated microphone and its correlation with drivers' impression," in *Proc. Audio Engineering Society 117th Convention*, Audio Engineering Society, San Francisco, CA, Paper 6307 (2004).
32. N. Samardzic and C. Novak. "Sound source signal parameters in vehicles for determining Speech Transmission Index," *J. Audio Eng. Soc.*, **59**, 735-744 (2011).
33. M. Nilsson, M., S.D. Soli, and J.A. Sullivan. "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 (1994).
34. J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, Mass, pp. 1-427 (1997).
35. N. Samardzic and C. Novak. "The band importance function in the evaluation of the speech intelligibility index at the speech reception threshold within a simulated driving environment," in *Proc. SAE Noise and Vibration Conference*, SAE, Paper 2013-01-1953 (2013).
36. N. Samardzic and C. Novak. "A novel method for in-vehicle speech intelligibility evaluation and statistical variability analysis," *Int. J. Vehicle Noise. Vib.*, **11**, 114-132 (2015).
37. N. Samardzic. "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. Sys.*, **7**, 434-438 (2014).
38. K. Genuit. "The sound quality of vehicle interior noise: a challenge for the NVH-engineers," *Int. J. Vehicle Noise. Vib.*, **1**, 158-168 (2004).
39. A.W. Bronkhorst and R. Plomp. "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *J. Acoust. Soc. Am.* **83**, 1508-1516 (1988).

40. B.C.J. Moore. *An Introduction to the Psychology of Hearing, 6th Ed.*, Brill, Leiden, The Netherlands, pp. 1-441 (2012).
41. B.C.J. Moore. "Development and current status of the "Cambridge" loudness models," *Trends Hear.*, **18**, 1-29 (2014).
42. E. Zwicker and U.T. Zwicker. "Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution," *J. Acoust. Soc. Am.*, **89**, 756-764 (1991).
43. B. Scharf. "Dichotic summation of loudness," *J. Acoust. Soc. Am.* **45**, 1193-1205 (1969).
44. B.R. Glasberg and B.C.J. Moore. "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.*, **47**, 103-138 (1990).
45. B.R. Glasberg and B.C.J. Moore. "A model of loudness applicable to time-varying sounds," *J. Audio Eng. Soc.*, **50**, 331-342 (2002).
46. B.C.J. Moore, M. Jervis, M., L. Harries, and J. Schlittenlacher. "Testing and refining a loudness model for time-varying sounds incorporating binaural inhibition," *J. Acoust. Soc. Am.*, **143**, 1504-1513 (2018).
47. B.R. Glasberg and B.C.J. Moore. "Development and evaluation of a model for predicting the audibility of time-varying sounds in the presence of background sounds," *J. Audio Eng. Soc.*, **53**, 906-918 (2005).

Table 1—The speech intelligibility metrics SII and STI calculated using the measured spectrum of speech with an overall level of 60 dBA, for rough road surface conditions simulated using a dynamometer with a simulated vehicle speed of 50 km/h and 100 km/h, shown separately for the left and right ears. The listener was always located at the front-left (driver) position.

Vehicle speed, km/hr	Talker location	SII Left	SII Right	STI Left	STI Right
50	Front right	0.62	0.78	0.66	0.86
100	Front right	0.40	0.60	0.43	0.66
50	Rear right	0.66	0.71	0.63	0.72
100	Rear right	0.43	0.50	0.40	0.50
50	Rear left	0.59	0.63	0.63	0.68
100	Rear left	0.37	0.42	0.40	0.46

Table 2—BLR, SII, and STI values at the SRT for different talker locations and background noise conditions. The listener was always located at the front-left (driver) position. The metrics for the SII and STI are for the the ear of the listener that was closer to the talker. The changes in the SNR in dB (applied to both ears) required to obtain BLR, SII, and STI values equal to the average value for a given metric are also shown. The bottom line shows the coefficient of variation (SD/mean).

Situation	Vehicle speed, km/hr	Talker location	BLR (sone/sone)	Change in SNR to get average BLR (dB)	SII	Change in SNR to get average SII (dB)	STI	Change in SNR to get average STI (dB)
	50	Front right	0.68	–0.8	0.49	–3.2	0.50	–3.4
	100	Front right	0.61	0.8	0.42	–0.4	0.40	–0.3
Dynamometer	50	Rear right	0.74	–2.1	0.47	–2.3	0.46	–2.3
	100	Rear right	0.66	–0.5	0.33	1.8	0.32	1.9
	50	Rear left	0.76	–2.6	0.49	–3.0	0.50	–3.2
	100	Rear left	0.64	0.0	0.36	1.1	0.36	0.9
	50	Front right	0.59	1.4	0.39	0.2	0.37	0.5
	100	Front right	0.61	0.4	0.44	–0.9	0.43	–0.7
On road	50	Rear right	0.63	0.2	0.33	2.0	0.32	2.3
	100	Rear right	0.59	1.3	0.30	2.9	0.29	3.0
	50	Rear left	0.64	0.0	0.36	0.9	0.36	0.8
	100	Rear left	0.57	1.8	0.36	1.2	0.36	0.9
Mean			0.643		0.394		0.390	
SD			0.059	1.3	0.066	2.0	0.069	2.1
Coefficient of variation			0.091		0.168		0.178	

Figure Captions

Fig. 1—Vehicle on a dynamometer in a semi-anechoic chamber for speech and noise measurements for a typical configuration, as used in [30]; the driver was the listener and the talker was the front passenger. Separate HATS systems were used to simulate the listener and the talker.

*Fig. 2— **(left)** A human subject driving the simulator while listening to the HINT sentences. **(right)** The HATS fitted with the headphones so as to check the calibration of the stimuli delivered to the listener.*

Fig. 1—Vehicle on a dynamometer in a semi-anechoic chamber for speech and noise measurements for a typical configuration; the driver was the listener and the talker was the front passenger. Separate HATS systems were used to simulate the listener and the talker.

*Fig. 2— **(left)** A human subject driving the simulator while listening to the HINT sentences. **(right)** The HATS fitted with the headphones so as to check the calibration of the stimuli delivered to the listener.*