

The Applicability of the Objective Speech Intelligibility Metrics for Vehicle Interior Speech Intelligibility Evaluation, Considering Different Listening Configurations and Background Noise Spectra

Nikolina Samardzic

ABSTRACT

Values of the speech intelligibility index (SII) were found to be different for the same speech intelligibility performance measured in an acoustic perception jury test with 35 human subjects and different background noise spectra. Using a novel method for in-vehicle speech intelligibility evaluation, the human subjects were tested using the hearing-in-noise-test (HINT) in a simulated driving environment. A variety of driving and listening conditions were used to obtain 50% speech intelligibility score at the sentence Speech Reception Threshold (sSRT). In previous studies, the band importance function for 'average speech' was used for SII calculations since the band importance function for the HINT is unavailable in the SII ANSI S3.5-1997 standard. In this study, the HINT jury test measurements from a variety of background noise spectra and listening configurations of talker and listener are used in an effort to obtain a band importance function for the HINT, to potentially correlate the calculated SII scores with the measured speech intelligibility scores. If a mathematical solution for such a problem is not found, then the measured speech intelligibility at the sSRT, considering different background noise spectra, is not, as expected, associated with a particular SII score; for example, 0.5. As a result, SII at sSRT may not be a suitable method for future in-vehicle speech intelligibility evaluation or target-setting, as it is sensitive to the changes in (speech) signal to (background) noise ratio and/or background noise spectrum shape.

CITATION: Samardzic, N., "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.* 7(1):2014, doi:10.4271/2014-01-9126.

INTRODUCTION

In today's society, a vehicle interior has become a common communication environment. The effectiveness of in-vehicle speech communication can be a good indicator of the perception of the overall vehicle quality and customer satisfaction. A complete and accurate assessment of speech intelligibility in the context of a vehicle interior listening environment is critical. Currently available objective speech intelligibility metrics - Articulation Index (AI), Speech Transmission Index (STI) and Speech Intelligibility Index (SII) - do not account in their procedures for essential parameters needed for this type of assessment. These parameters include the directivity and the distance of the talker with respect to the listener, binaural listening, hearing profile of the listener, vocal effort, and multisensory hearing.

The effectiveness of in-vehicle application of the available objective metrics was investigated in a series of studies to reveal their shortcomings, including a wide range of scores resulting from each of the metrics for a given measurement configuration and vehicle operating condition [1, 2, 3, 4]. In addition, it was found that the nature of a possible correlation between the scores obtained from each metric is unknown.

The metrics and the subjective perception of speech intelligibility using, for example, the same speech material have not been compared in literature. As a result, an alternative method for speech intelligibility evaluation was proposed for use in the automotive industry by utilizing a virtual reality driving environment for ultimately setting targets, including the associated statistical variability, for future in-vehicle speech intelligibility evaluation [5]. The Speech Intelligibility Index (SII) was evaluated at the sentence Speech Reception Threshold (sSRT) for various listening situations and hearing profiles using acoustic perception jury testing and a variety of talker and listener configurations and background noise [6]. In addition, in the same study, the effect of individual sources and transfer paths of sound in an operating vehicle to the vehicle interior sound, specifically their effect on speech intelligibility, was quantified, in the framework of the newly developed speech intelligibility evaluation method.

Lastly, as an example of the significance of speech intelligibility evaluation in the context of an applicable listening environment, as indicated in this research, it was found that the jury test participants required on average an approximate 3 dB increase in sound pressure level of speech material with

driving and listening compared to when just listening, for an equivalent speech intelligibility performance and the same listening task [5].

In previous studies it was found that the SII at sSRT is sensitive to the changes in SNR; the SII scores were lower for the on-road background noise measurements [6], compared to the dynamometer background noise measurements [5], for the same speech intelligibility performance, using the same vehicle, driving and listening tasks and listeners' hearing profile. In both studies [5,6], the band importance function for average speech, was assumed to be adequate for the hearing-in-noise test (HINT), and used for the SII calculations, as the band importance function for the HINT has not been developed for ANSI-S3.5-1997. The rationale was that, according to the ANSI S3.5-1997, the band importance function for average speech provides accurate predictions across different communication situations and for communications situations where contextual, linguistic, semantic and syntactic constraints vary within a situation. For constant background noise conditions, it was found that the SII at sSRT is relatively insensitive to the band importance function [7]; in this study, on-road background noise data was used to create a driving simulator model for acoustic perception jury testing [6].

The fact that SII is sensitive to SNR, given the same speech intelligibility performance, may suggest that the assumed average speech band importance function is not appropriate for the HINT. If this is the case, a band importance function yielding the same SII score, regardless of the SNR, background noise level or spectrum shape, corresponding to the same speech intelligibility performance (i.e. sSRT) from an acoustic perception jury test, could be calculated. Otherwise, there is no unique SII score associated with the sSRT, and the SII metric could not be used to characterize sSRT for the listening situation investigated in this research. Current study attempts to investigate this in more detail.

NOVEL METHOD FOR SPEECH INTELLIGIBILITY EVALUATION

The common objective speech intelligibility metrics require that a particular level of vocal effort be specified in their analysis regardless of the evaluation method used. Ultimately, the metrics are evaluated in terms of the scores ranging from zero to one and interpreted based on the subjective descriptions of scores associated with a particular evaluation method or a standard, if such descriptions are available. An alternative, as suggested in previous work, to using a particular vocal effort in the evaluation and later interpretation of objective speech intelligibility scores, is to consider the sentence speech reception threshold thus eliminating the need for the subjective interpretation of scores by, instead, characterizing the human speech hearing ability in presence of noise. The sentence speech reception threshold (sSRT) is defined as "the minimum hearing level for speech at which an individual can recognize 50% of the speech material" [8]. Utilizing the sSRT eliminates

the need for the subjective interpretation of scores by, instead, characterizing the human speech hearing ability in the presence of noise. For sentence-type tests typically used to measure sSRT, there is a point (a sound pressure level) after which the speech is understood in its entirety. Below that point, it is not fully understood. This speech intelligibility evaluation method is easy to understand, simple and universally applicable. In this study, the HINT is used for obtaining sSRT measurements using acoustic perception jury testing.

HINT MEASUREMENT PROCEDURE

The jury test participants in the first acoustic perception study [5], using vehicle dynamometer background noise measurements, consisted of 30 participants, 18 males and 12 females between ages of 19 and 30. The average age of the participants was 24 (standard deviation 3.2). All of the participants had normal hearing. The jury test participants in the second acoustic perception study [6], using on-road background noise measurements, consisted of 10 participants, 9 males and 1 female. In this group, 4 male participants were hearing impaired, their ages ranging from 22 to 67. Their audiograms with hearing threshold information were provided by the participants and used in subsequent SII calculations. The remaining normal hearing participants' average age was 25 (standard deviation 2.8). In the current study, only the data from normal hearing participants is considered for calculations of the band audibility function.

The HINT adaptive measurement procedure consisted of increasing the gain in 2 dB increments from a level below the expected sSRT, until the sentence was repeated correctly. The gain was then decreased by 2 dB for the second sentence. If the second sentence was then repeated correctly, the gain was again decreased by 2 dB; otherwise, it was increased by 2 dB for the next sentence. The same procedure was followed for the remaining 20 sentences, as the gain was adjusted by +/- 2 dB steps. The sSRT was obtained as the average of the presentation levels from sentence 5 to sentence 21; including the presentation levels of both the correctly and the incorrectly repeated sentences.

For both jury tests, twenty HINT sentences were presented via headphones (corrected for distortion) at intervals of approximately 20 seconds for each of the six measurement configurations considered. Since the recognition of speech material becomes less difficult as it is repeated or reused, each test condition in this study included 20 unique HINT sentences. The duration of the simulated driving test using the HINT and the driving simulator was approximately one hour. The participants were first asked to drive at 50 km/h during which the HINT sentences representing the talker in the passenger front location were played through the headphones. These sentences were played simultaneously with the background noise associated with a particular driving condition (gear, throttle position, engine revolutions per minute). A short break was subsequently provided for the participants before starting a 100 km/h test for the same talker and listener configuration.