

Automotive Speech Intelligibility Measurements

Gordon Ebbitt and Todd Remtema
Toyota Technical Center USA Inc.

ABSTRACT

Speech communication from the front seat to the rear seat in a passenger vehicle can be difficult. This is particularly true in a vehicle with an acoustically absorptive interior. Speech Transmission Index (STI) measurements can quantify the speech intelligibility, but they require specialized signal processing. The STI calculations can be simplified if it is assumed that reverberation and echoes play an insignificant role in an automobile. A simplification of a STI measurement is described that uses a stationary reference speech signal from a talker mannequin in the driver's seat to create a signal at the rear passenger positions. On-road noise measurements are used for the noise level and the calculated signal to noise ratio is used to calculate a simplified STI value that tracks closely to a full implementation of the STI method for sedans. In fact, this method is very similar to the techniques described in the Articulation Index (AI) and Speech Interference Index (SII) standards. Those standards provide for the use of a talker mannequin, though that is rarely done. The use of all three measurement techniques will be considered.

CITATION: Ebbitt, G. and Remtema, T., "Automotive Speech Intelligibility Measurements," *SAE Int. J. Passeng. Cars - Mech. Syst.* 8(3):2015, doi:10.4271/2015-01-2337.

INTRODUCTION

In an attempt to make vehicles quieter it is tempting to add a great deal of acoustic absorption to the interior. This, however, may cause the speech intelligibility to become worse. For many years the ability to communicate easily in a vehicle has been considered a key attribute contributing to the perception of a comfortable acoustic environment and a high quality vehicle. The ability to hear and understand speech is obviously important for all vehicle occupants, but can be a particular issue when the driver communicates with passengers in the rear of the vehicle. Vehicles with acoustically absorptive interiors can be particularly problematic and vehicles with third row seats provide even bigger challenges.

The three most common measurements to assess intelligibility are Articulation Index (AI), Speech Intelligibility Index (SII), and Speech Transmission Index (STI). Each of these measures provides a method for generating a signal at the talker's position, measuring the response at the listener's position, and judging the ability to understand speech based on the speech to noise level at the listener position.

This, however, is not the way the Articulation Index is used in the automotive industry. Articulation Index is generally not used to calculate an actual speech intelligibility rating, but rather a "speech weighted" noise level. The Articulation Index is calculated by computing the speech to noise level, but an idealized speech spectrum in the Articulation Index standard is generally used in place of actually measuring the received speech level at the listener's position. Though the Speech Intelligibility Index is a more recent measure and

contains updates and refinements to the Articulation Index standard, it has not been widely adopted. Both of these measures will be considered in more detail below.

In recent years there has been quite a bit of interest in using the Speech Transmission Index to gauge speech intelligibility. This method cannot be implemented without simulating talker levels. STI accounts for the drop in level as the sound travels from the talker to the listener, the background noise level at the listener's position, and reverberation and echoes in the environment. STI requires the measurement of a modulation transfer function between the talker and listener which accounts for the effects of reverberation, echoes, and background noise. While the AI and SII computations are relatively simple, the STI method is somewhat more complicated. It will be shown that the STI method yields results which are similar to a simpler AI/SII style measurements in a vehicle where echoes and reverberation are not significant.

All of these measures, AI, SII, and STI, assume a single receiving microphone. For automotive measurements, however, there can be large differences between measurements at the inboard and outboard ears of a rear passenger. In general, the intelligibility will be much higher for the inboard ear. This is because the received talker level is higher at the inboard ear and the vehicle noise is somewhat lower at this location. Though there have been studies of binaural STI in larger architectural spaces, no work has been done to investigate binaural effects in automobiles. That will be considered.

ARTICULATION INDEX AND SPEECH INTELLIGIBILITY INDEX

Procedures to quantify articulation index have existed since, at least, 1929 [1] when the term was used to describe a juried method with talkers, listeners, and lists of words. This method was primarily intended for communications systems (telephone and radio), but its use for other applications was also considered.

An objective measurement technique was standardized in ANSI S3.5-1969, American National Standard Methods for the Calculation of the Articulation Index [2]. That standardized measurement technique is based on an assessment of the ratio of measured or “idealized” speech and the background noise level at the listener's position. The 1969 standard has been superseded by ANSI S3.5-1997, American National Standard Methods for the Calculation of the Speech Intelligibility Index [3]. The 1997 standard provides a method which is similar in principle to the earlier standard, but contains enough differences that the authors choose a new name for the measured quantity. Some of the many differences include the use of different standardized talker levels, an expanded frequency range, and updated band importance factors. Note that the original 1969 standard had a single talker level. The updated standard has four talker levels: “normal”, “raised”, “loud”, and “shouting”.

Both the 1969 and 1997 standards rely on a computation of the speech to background noise difference at the listener's position. This ratio is computed in bands: octave, 1/3-octave, or bands that contribute equally to the intelligibility of speech (in addition, the 1997 standard includes critical bands). The speech to noise difference in these bands is multiplied by a band importance factor and the results are summed over all the frequency bands. A speech to noise difference of 30 dB at all bands will result in an Articulation Index of 1 while a speech to noise difference of 0 at each band will result in an Articulation Index of 0. To keep the Articulation Index from exceeding 1 or becoming less than 0, speech to noise differences are capped at a maximum of 30 dB and a minimum of 0 dB. This range defines the limits from perfect speech intelligibility to no intelligibility at all.

An example AI calculation from ANSI S3.5-1969 is shown in Table 1. The peak speech levels and the background noise from this example are shown in Figure 1. Note that the RMS speech levels are assumed to be 12 dB lower than the peak levels.

The differences between AI and SII are not large, but can be significant in cases where the Articulation Index is being used as a vehicle development target. In one study, vehicle level measurements indicated that the SII values are about 3% to 4% less than the AI values during a vehicle acceleration [4].

Note that many automotive OEM's and suppliers have created their own talker levels and band importance functions. An additional complication is that the newer ANSI Speech Intelligibility Index standard has not been widely adopted in the automotive industry. Though ANSI S3.5-1969 (the AI standard) has been out of print since it was replaced by ANSI S3.5-1997 (the SII standard), AI is still more widely used as compared to SII.

Table 1. This is an example of the Articulation Index calculation from ANSI S3.5-1969. The speech spectrum consists of peak levels based on the assumption of a 65 dB long term RMS idealized speech level at the listener's position.

Freq	Peak Speech Level (RMS + 12dB) taken from Idealized Speech Spectrum (Long term RMS = 65 dB)	Background Noise (dB)	Speech Peaks minus Background Noise (dB)	Weighting Factor	
200	67.0	63.0	4	0.0004	0.0016
250	68.0	58.0	10	0.0010	0.0100
315	69.0	56.0	13	0.0010	0.0130
400	70.0	46.0	24	0.0014	0.0336
500	68.5	42.5	26	0.0014	0.0364
630	66.5	40.5	26	0.0020	0.0520
800	65.0	41.0	24	0.0020	0.0480
1000	64.0	43.0	21	0.0024	0.0504
1250	62.0	44.0	18	0.0030	0.0540
1600	60.5	42.5	18	0.0037	0.0666
2000	59.5	44.5	15	0.0037	0.0555
2500	58.0	43.0	15	0.0034	0.0510
3150	56.0	50.0	6	0.0034	0.0204
4000	53.0	45.0	8	0.0024	0.0192
5000	51.0	39.0	12	0.0020	0.0240
				AI =	0.5357

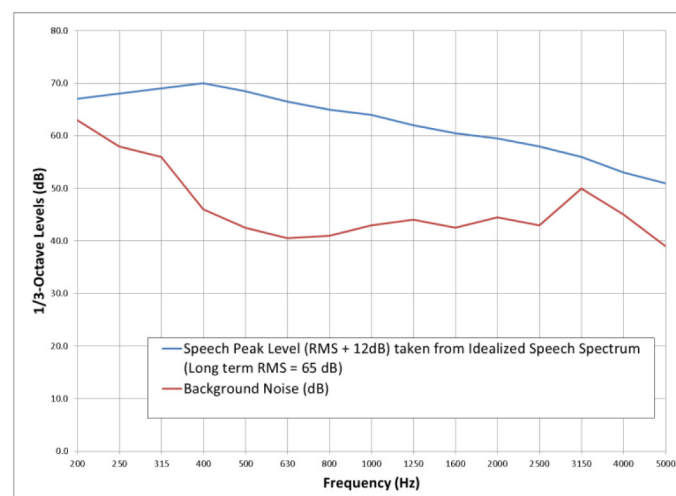


Figure 1. Peak speech level and background noise from an example calculation in ANSI S3.5-1969. The speech levels are the peak levels in a speech signal with a long term RMS average of 65 dB.

As already mentioned, the AI and SII standards provide for a measurement of the talker's speech level at the listener's position, this is not the way these standards are generally used for automotive work. It can be cumbersome (or, at least, an extra measurement step) to mount a speaker at the talker's position and measure the levels at the listener position. For this reason, the standardized speech levels in the standards are often used as the signal in the signal to noise difference calculations. The resultant AI and SII values are, therefore, a “speech intelligibility weighted” noise level rather than an actual assessment of the speech to noise ratio at the listener's position.

SPEECH TRANSMISSION INDEX

The Speech Transmission Index was developed by Steeneken and Houtgast in the late 1960's to measure speech intelligibility in VHF radio systems and is now standardized by the International Electrotechnical Commission (IEC) [5]. Speech Transmission Index measurements are similar to AI and SII measurements in that they require the generation of a signal at the talker's position and a measurement of that signal at the listener's position. As with AI and SII, STI measurements use a test signal with a frequency spectrum equal to the long term frequency spectrum of speech. STI measurements differ in that they use a modulated noise source. For a male speaker octave band noise from 125 Hz to 8000 Hz (7 bands) is used and each octave band is modulated with a periodic signal in such a way that the intensity envelope is modulated sinusoidal. For a female speaker the 125 Hz octave-band is not used. The modulation frequencies range from 0.63 Hz to 12.5 Hz in 1/3-octave steps (14 frequencies). This is a total of 98 measurements for a male speaker.

The measured signal is influenced by its travel through the acoustic environment (including the drop in level as well as the influence of reverberation and echoes) and by the background noise at the listener's position. The STI technique is fundamentally different as compared to the AI or SII methods in that it assesses the temporal degradation of the speech signal as well as the speech to background noise ratio.

The change in the modulation is the modulation transfer function, $m(F_{k,f})$, where

$$m(F_{k,f}) = \frac{m_o}{m_i} \quad (1)$$

and m_o is the modulation of the signal at the listener's position, m_i is the modulation of the signal at the talker's position, k indicates one of the 7 octave bands (125 Hz to 8k Hz) and f designates one of the 14 modulation frequencies (.63 Hz to 12.5Hz). This is determined for all 98 cells as shown in Figure 2. Note that Equation (1) and all the following equations related to STI measurements can be found in the STI standard [5].

Though STI could be determined by measuring all 98 modulated sounds, STI measurements are generally implemented by measuring the frequency response function from the talker to listener position and calculating the impulse response function. The modulated sound at the speaker's position can then be determined from the impulse response.

The modulation transfer function can be interpreted as an apparent signal to noise ratio (SNR). This signal to noise ratio includes factors related not only to the difference between the speech and background noise levels, but also the effects of reverberation and echoes.

$$SNR_{k,f} = 10 \log \left(\frac{m(F_{k,f})}{1 - m(F_{k,f})} \right) \quad (2)$$

Note that the apparent SNR in the STI calculations is a little different as compared to the signal to noise ratio as it is used in the AI and SII calculations. For the AI and SII calculations it is just the difference between the speech level at the listener's position and the background noise. For AI and SII calculations a signal to noise ratio of 30 dB would indicate 100% intelligibility and a signal to noise ratio of 0dB would lead to 0% intelligibility. The apparent SNR in the STI calculations is similar to the signal to noise ratio in AI and SII calculations in that there is a linear relationship between the SNR and intelligibility. The apparent SNR in the STI calculation, however, spans a range from $-\infty$ to $+\infty$ with an apparent SNR of 0 dB corresponding to a modulation of 0.5. The apparent SNR is truncated outside the range -15 to $+15$ which corresponds to intelligibility scores from 0 to 1. Differences in the way the SNR is calculated are accounted for by the Transmission Index which is described below.

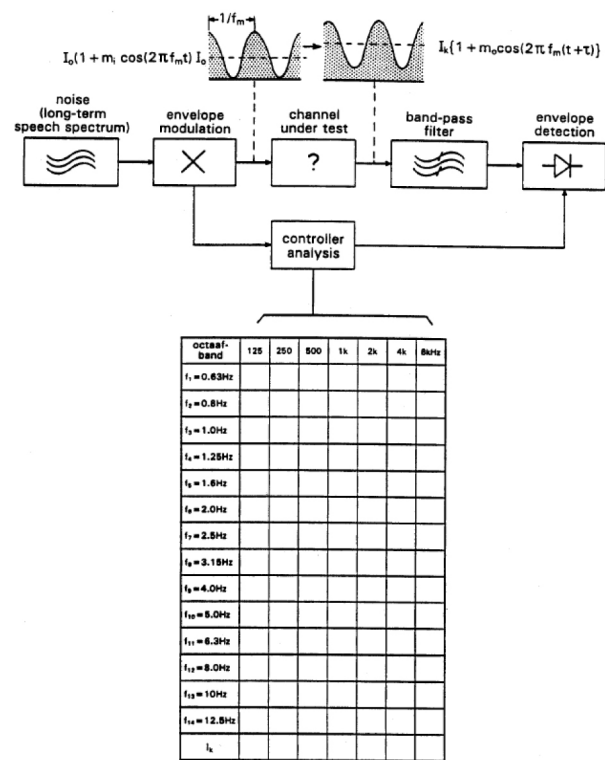


Figure 2. A schematic representation of the STI measurement process. Octave band sound from 125 Hz to 8 kHz is modulated at 14 frequencies from 0.63Hz to 12Hz and the modulation transfer function is measured from the source to receiver positions [5].

This apparent SNR is calculated for each of the 98 cells in Figure 2. The SNR is converted to a transmission index $TI_{k,f}$

$$MTI_f = \frac{SNR_{k,f} + \text{shift}}{\text{range}} \quad (3)$$

where shift = 15 dB and range = 30 dB.

The modulation transfer index, MTI_k , is the average of all transmission indices for a given octave band frequency.