Effect of Human Speaker's Head Rotation on Speech Transmission Index in Vehicle Sound Environment

Le Yu[1], Guangzheng Yu[1,*], Qinglin Meng[1,†]
[1] *Acoustic Laboratory, School of Physics and Optoelectronics, South China Univ. of Tech., Guangzhou, 510641, China*
E-mail: * scgzyu@scut.edu.cn; † mengqinglin@scut.edu.cn

Abstract **— The speech transmission index (STI) is one of most common objective speech intelligibility metrics. The value of STI metric ranges from 0, indicating unintelligible speech, to 1, indicating excellent speech intelligibility, which could be influenced by signal to noise ratio (SNR), reverberation (or echo), and driving environment. In addition, a passenger talker could actually rotate his / her head to the driver when he / her talks to driver, which could induce the influence of speaker directivity on the STI. In current work, we measured several groups of binaural impulse responses on a human subject, under different orientations of a human speaker, and then calculate the corresponding STI values. Finally, we analyze the variation of STI caused by the human speaker's orientations, and give some advices on the measurement and evaluation method for STI in vehicle.**

I. Introduction

The sound field in an automobile cabin is very complex because of the small space, irregular shape, and arrangement of sound absorbing material. In addition, the existing external noise and interior noise result in a challenging communication environment inside vehicle. The speech intelligibility is an important research field for communication environment, and the complex sound field could directly result in the loss of speech intelligibility. Researchers always pay attention to the factors of influencing the loss of speech intelligibility between the talker and the listener and dedicate to develop the physical metrics to predict the speech intelligibility in vehicle, which is of great significance.

In 1992, E. Parizet [1] investigated on the calculation of the articulation index (AI) in the car and compared the influence of different researchers on the definition of frequency band weight index on the AI. In the same period, the influence of two types of speaker systems on AI was also analyzed, the author pointed out that AI of the omnidirectional speaker system in-vehicle is similar to the AI in the free field, and the interior of vehicle equipped with a directivity speaker system helps increase AI [2]. N. Samardzic [3] studied the problem of loss of speech intelligibility in various vehicle interior spaces under different operating conditions and road conditions. The three metrics include AI, speech intelligibility index (SII) and speech transmission index (STI), which can be used for automobile interior speech intelligibility assessment, and were introduced and tested in actual vehicles. During the time, the author pointed out that the AI approach involves only measurement of background noise and does not capture the effects of the vehicle cabin surface, the characteristics of speech signal, the distances between talkers and listeners, or the directivity of the talkers, and thus couldn't be a better predictor of speech intelligibility inside vehicle [4].

The STI is one of the most common objective metrics to predict speech intelligibility by considering noise, speech signal and room acoustic properties, and it has successfully been applied to many room acoustic environments, such as classroom [5], office, and laboratory [6]. Especially, STI has recently been used in the application of the automobile interior speech transmission in terms of intelligibility [7-9].

However, all of these researches were based on the artificial head and artificial speaker (or talker), in which head rotation of human speaker hasn't been concerned. In fact, if a speaker seats in the front compartment and faces to the driver, the listener could obtain better STI because of speaker's radiation directivity. Therefore, in this study, we measured several groups of binaural impulse responses on a human listener by changing the rotations of artificial talker, so as to simulate the human speaker's head rotation. Then, we generate the binaural sound pressure by convoluting binaural impulse response with speech signal under the same background noise condition. At last we calculate the value of STI by using the convolved signals and analyze the influence of the head rotation of the speaker on STI in a vehicle.

II. Theory

A. STI calculation

The STI metric was proposed by Steeneken and Houtgast [10] in 1973, and was shown to be well correlated to speech intelligibility. Then the STI calculation program was standardized as described in IEC 60268-16:2003 [11]. For STI calculation, an input signal with a sinusoidal intensity modulation $I_i(\mathrm{t}) = I_i(1 + m_i \cos 2\pi F t)$, I_i is sound intensity amplitude, F is the modulation frequency, and the output signal $I_o(\mathrm{t}) = I_o(1 + m_o \cos 2\pi F(t+\tau))$, τ is the phase difference, respectively, are firstly used in the calculation of Modulation Transfer Function (MTF, denoted by symbol '*M*'), defined as the ratio of the modulation index of source signal, m_i and the modulation index of receiver signal, m_o. M is distributed in 14 modulation frequencies, *F*, at one third octave intervals, ranging from 0.63 Hz to 12.50 Hz and 7 octave band frequencies, *f*, ranging from 1.25 kHz to 8 kHz. Then we will get the $\{m_{f,F}\}$ matrix with size of 14 lines and 7 columns, and the apparent signal to noise ratio ($SNR_{f,F}$) is defined as:

$$SNR_{f,F} = 10\log_{10}\left(\frac{m_{f,F}}{1-m_{f,F}}\right) \quad (1)$$

The $SNR_{f,F}$ is limited in the range of -15dB to +15dB, indicating that STI linearly variation from 0 to 1.

$$SNR_{f,F} = \begin{cases} -15, if\ SNR_{f,F} < -15 \\ SNR_{f,F} \\ 15, if\ SNR_{f,F} > 15 \end{cases} \quad (2)$$

Equation (2) means that if SNR is less (more) than -15dB (15 dB), the STI should be assumed to be absolutely unaccepted (accepted) in the cabin, respectively. Then, the transmission index ($TI_{f,F}$) is defined as:

$$TI_{f,F} = \frac{SNR_{f,F}+15}{30} \quad (3)$$

The modulation transmission index (MTI_f) can be obtained from $TI_{f,F}$:

$$MTI_f = \frac{1}{14}\sum_{F=1}^{14} TI_{f,F} \quad (4)$$

The revised STI is then calculated as the weighted sum of MTI_f:

$$STI_r = \sum_{n=1}^{7}\alpha_n MTI_n - \sum_{n=1}^{6}\beta_n\sqrt{MTI_n \cdot MTI_{n+1}} \quad (5)$$

The α_n and β_n are the revised weight factors of male speech signal as shown in Table 1.

In order to reduce the measurement and increase the repeatability of impulse response, the $m_{f,F}$ is usually obtained from the noise-free impulse response h_f, speech signal level L_{signal} and background noise level L_{noise} [12].

$$m_{f,F} = \frac{\int_0^\infty h_f^2(\tau)\exp(-j2\pi F\tau)\,d\tau}{\int_0^\infty h_f^2(\tau)d\tau} \cdot \frac{1}{1+10^{\frac{L_{noise}-L_{signal}}{10}}} \quad (6)$$

The $m_{f,F}$ is the modulation transfer function at modulation frequency F, $h_f(\tau)$ is the impulse response octave band filtered at carrier *f*. L_{noise} is defined as the binaural noise signal sound pressure level at octave band frequency *f*. L_{signal} is the binaural speech signal sound pressure level at octave band frequency *f*.

Table 1. Weighting factors for male speech

Octave Band (Hz)	*125*	*250*	*500*	*1000*	*2000*	*4000*	*8000*
α	0.085	0.127	0.23	0.233	0.309	0.224	0.173
β	0.085	0.078	0.065	0.011	0.047	0.095	—

In this paper, the noise-free binaural impulse responses ($h(\tau)$) were firstly measured in a quiet vehicle environment, then $h_f(\tau)$ could be calculated by filtering at 7 octave band frequencies *f*. Binaural noise signals were generated by the white noise signal. Binaural speech signals were calculated by convoluting the speech signal with noise-free binaural impulse responses, while the speech signal was obtained by generating a pink noise signal, and then filtering and adjusting the spectrum according to the GB/T 7347-1987, *The standard spectrum of Chinese speech* [13], L_{signal} is the binaural speech level at 7 octave band frequencies *f*.

B. *Experiment equipment*

As aforementioned, $h_f(\tau)$, L_{noise} and L_{signal} should be obtained through measurement of the binaural impulse responses. Therefore, binaural impulse responses should be first obtained.

In this experiment, the binaural impulse responses were measured in an automobile cabin (Volkswagen Sagitar) in our quiet campus environment. A human subject sitting in the driver location was considered as the receiver (listener) with his head forward. A Mouth Simulator (Bruel & Kjaer Type 4227 Unit) was used as the talker (source), with rotation angle ranged from 0° to 180°, 45° intervals, in the horizontal plane.

C. *Signal processing*

A maximum length sequence (MLS) signal was then played by the mouth simulator at all angles, increasing the gain of the MLS signal, in order to maximize the signal to noise ratio (SNR). A pair of miniature microphones was placed in the ear canals of the human object, picking up the binaural signal for all the rotation angles of the talker considered in the study. Then, the binaural impulse responses were captured by the cyclic cross-correlation of binaural signals and MLS signal as shown in Fig. 1.

Speech signal was obtained by generating a pink noise signal, filtering and adjusting the spectrum according to the GB/T 7347-1987, then the binaural speech signals were captured by convoluting the speech signal with binaural impulse responses, and then, the octave band sound pressure level of the binaural speech signals (L_{signal}) were calculated by filtering at specific frequencies *f*. In this study, white noise was acted as the binaural noise signals. The octave band sound pressure level of binaural noise signals (L_{noise}) also could be calculated through the above method.

The modulation transfer metrics were obtained using the noise-free impulse response method, then, the revised STI (STI_r) can be calculated.

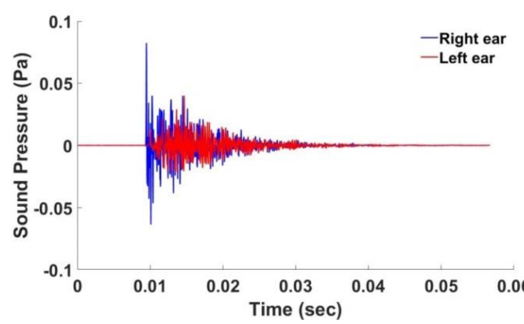

Fig. 1 The binaural impulse response of listener with the listener and the talker face straight ahead, illustrating the short reverberation time.

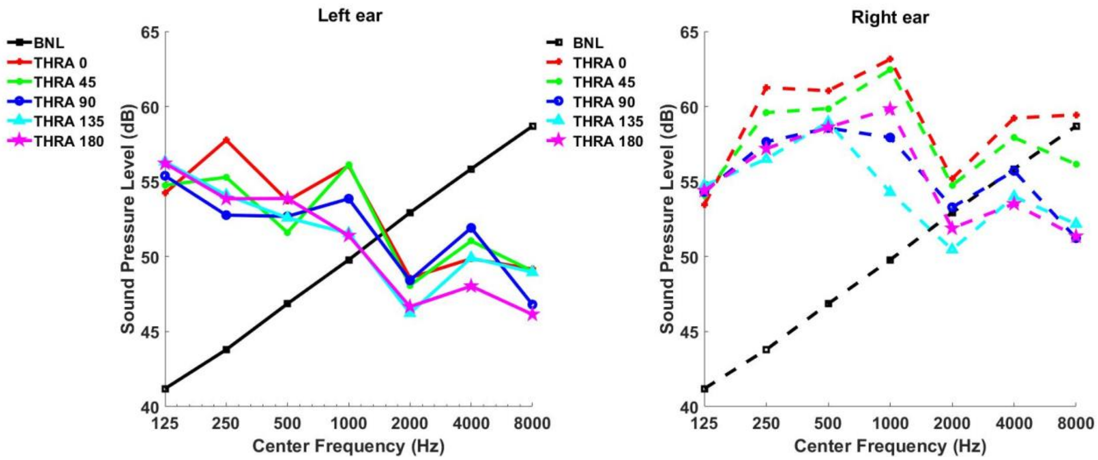

Fig. 2 Binaural speech signal and noise signal octave band sound pressure level of listener, with the listener forward, the talker head rotation angle (THRA) changing from 0° to 180°, 45° intervals. White noise acts as the background noise signal (BNL).

III. Results and discussion

The binaural STI_r are calculated at 5 directivities of the Mouth Simulator, the results provide a relatively complete analysis of factors influencing the loss of speech intelligibility, such as the influences of directivity of the speaker and surface reflection of the cabin, etc. The maximal value of STI, 1.0, signifies that all speech cues can reach the listener, while its minimum value, 0, signifies that no speech cues is available to the listener. The higher value of STI corresponds to the smaller loss of the speech intelligibility.

A. Data of Binaural Speech and Noise Signal Level

In this study, white noise was used as the binaural background noise. The binaural speech signals were obtained by convolution and the octave band calculation method to analyze the influences of head rotation of STI. This method considers the impact on speech within 7 frequency bands whose center frequency ranges from 125Hz to 8000Hz.

Fig. 2 illustrates the binaural octave band sound pressure level of speech and noise signal with the listener forward and the talker's head rotation angle changing from 0° to 180°, 45° intervals. The angle of 0° represents the left side of the talker, while 180° represents the right side of the talker. It's easy to notice that under different octave band, speech signal sound pressure level between right ear and left ear are different, the speech signal level of left ear is obviously lower than right ear because of the head effect. The binaural speech signal of the listener reaches maximum value when the talker's head rotation angle is 0°, and furthermore, with the angle increasing, the speech signal level gradually decreases at different octaves. When the angle increases to 180°, the speech signal level has a slight increasing. It is easy to explain that the window reflection on the right of the talker has a positive effect on the increasing of binaural octave band speech signal level. The sound pressure level of right ear at center frequency 1000 Hz is the largest, and then reach the minimum value at 2000 Hz, this because of the influence of the automobile interior acoustic characteristics and speech spectral characteristics. Compared with the right ear, the head rotation of the talker has less obvious effect on the speech signal level of the left ear. With the change of talker's head rotation, speech signal levels of left ear at all octave bands are similar. From the above analysis, we can deduce that the head of listener and directivity of talker has an import influence on difference of speech signal level between right ear and left ear in such a complex communication environment, and furthermore, could has an impact on STI.

B. Speech Intelligibility and STI

Speech transmission index (STI) was proposed as an important physical metric to predict speech intelligibility which has a great influence on vehicle sound quality. Fig. 3 shows the value of binaural STI of listener with all the head rotation angles of the talker. It indicates that the STI value of right ear is higher than that of left ear. For the angle of 0°, the STI value of (ipsilateral) right ear reaches maximal, 0.74, which corresponds to the great speech intelligibility in this work. With talker's head angle changing from 0° to 180°, the value of STI of (ipsilateral) right ear gets smaller gradually, but the value of STI of left ear has no apparent change. For

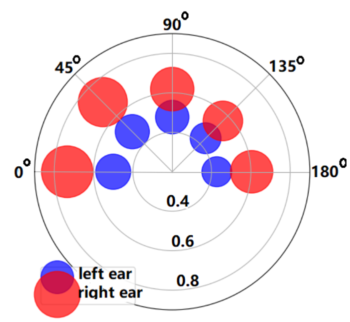

Fig. 3 The binaural value of revised STI, while listener's head face forward and talker's head rotation angles ranges from 0° to 180°, 45° intervals. The angle of 0° represents the left side of the talker and the angle of 180° represents the right side of the talker.

the angle of 180°, the STI value of right ear reaches 0.61, slightly higher than the value of at 135°, 0.56, which could be resulted from the right side window reflection to the speaker's signal, thereby enhancing the right ear STI of listener.

IV. CONCLUSION

In this study, we analyzed the speech intelligibility inside an automobile complex communication environment by using the STI parameter. The calculation program of the noise-free method has been used to acquire the revised STI under different rotation angle of the talker. Results show that the (ipsilateral) right ear STI value of listener is higher than the left ear, with the head rotation angle of the talker changing from 0° to 180°, the right ear STI value gets smaller gradually, while the left ear STI value has no apparent variation. For the talker's rotation angle of 0°, the STI value of the (ipsilateral) right ear reach maximum value, 0.74, indicating that in this case the loss of speech intelligibility is the lowest, speech should be clearer. In addition, The STI value of right ear at the talker's head rotation angle of 180° is slightly higher than that at 135° because of the window reflections. These results are beneficial to improve the normal speech communication and noise control inside a vehicle. Next, we will pay attention to the study of the improvement of STI calculation model considering the head rotation of both listener and speaker.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (grant nos. 11574090 and 11704129) and the Natural Science Foundation of Guangdong Province (grant no. 2018B030311025).

REFERENCES

[1] E.Parizet, "Influence of speech important function upon articulation index computation in cars," *Noise Control Engineering Journal*, vol.38, pp. 73-79, January 1992.

[2] E. Parizet, "Speech intelligibility mappings in a car compartment," *Int. J. of Vehicle Design*, vol. 14, 1993.

[3] Nikolina Samardzic, Colin Novak, "In-vehicle application of common speech intelligibility metrics," *Int. J. Vehicle Noise and Vibration*, vol. 7, No. 4, 2011.

[4] Nikolina Samardzic, Colin Novak, "In-vehicle speech intelligibility for different driving conditions using the Speech Transmission Index," *Noise Control Eng. J*, vol. 59, No. 4, July-Aug 2011.

[5] Jianxin Peng, Nanjie Yan, Dan Wang, "Chinese speech intelligibility and its relationship with the speech transmission index for children in elementary school classrooms," *J. Acoust. Soc. Am*, vol. 137, No. 1, pp. 85-93, January 2015.

[6] PeishengZhu, FangshuoMo, JianKang, "Relationship Between Chinese Speech Intelligibility and Speech Transmission Index Under Reproduced General Room Conditions," *Acta Acustica united with Acustica*, vol.100, No.5, pp.880-887(8), October 2014.

[7] F Bozzoli, A Farina, "Measurement of the speech intelligibility inside cars," in *Audio Engineering Society Convention 113*. Audio Engineering Society, 2002.

[8] M. Viktorovitch, Rieter Automotive AG, "Implementation of a new metric for assessing and optimizing the speech intelligibility inside cars," *Society of Automotive Engineers,* 2005-01-2478.

[9] G Ebbitt, T Remtema, "Automotive Speech Intelligibility Measurements," *Society of Automotive Engineers*, 2015-1-2337.

[10] H. J. M. Steeneken and T. Houtgast, "The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility," *J. Acoust. Soc. Am*, vol. 28, No. 2, pp.66-73(8), 1973.

[11] IEC 60268-16: 2003, Sound System Equipment – Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index, International Electrotechnical Commission, 2003.

[12] A.Farinaa, F.Bozzolib, P.Strasserc, "Comparative study of speech intelligibility inside cars," *Acta Neurologica Belgica*, vol. 103, No. 3, pp. 144-154, 2003.

[13] GB/T 7347-1987, the standard spectrum of Chinese speech, National Standard of China, 1987.