inter.noise
HAMBURG 2016

Speech intelligibility under in-car distant-talking environments

Mitsunori MIZUMACHI[1]; Shota TAKUMA[1]; Ikuyo OHSUGI[2]; Yasushi HAMADA[2]; Koichi NISHI[2]

[1] Kyushu Institute of Technology, Japan

[2] Mazda Motor Corporation, Japan

ABSTRACT

Speech intelligibility has been widely investigated for various purposes. Hands-free speech communication without any close-talking microphone becomes popular inside a vehicle for the safety reason. It is necessary to carefully install microphones so that speech signals are clearly captured under noisy in-vehicle environments. In this study, the relationship between the intelligibility of distant-talking speech and acoustical interferences is investigated inside car cabins. Target speech signals are prepared as the utterance of Japanese 4-digits random sequences. The acoustical interferences consist of exterior and interior noises. Difficulty in speech perception is defined by the speech reception threshold, which corresponds to the sound pressure level of speech at the 50 % correct score in an articulation test. It is suggested that the interior blower noise caused by an air-conditioner is the most dominant interferences in speech intelligibility. For further investigation, multiple regression analyses have been carried out using subjective indexes related to speech intelligibility. It is found that speech intelligibility can be modeled with thickness of speech, loudness of speech, and loudness of noise.

Keywords: Speech intelligibility and interference, speech levels, speech communication
I-INCE Classification of Subjects Number(s): 63.3

1. INTRODUCTION

It is difficult to exactly measure the intelligibility of distorted speech signals, although the prediction of speech intelligibility has been one of research issues in the fields of speech science and engineering [1]. The suitable intelligibility measure must be different depending on its application. Nevertheless, the articulation index [2], the speech intelligibility index [3], and the speech transmission index [4] are widely employed to quantify speech quality. This paper aims at measuring how comprehensible speech signals are in distant-talking conditions inside a car cabin. A driver should not hold a phone while driving, and is recommended to take the call using a hands-free speech communication device. Recent vehicles mostly equip with microphones for achieving hands-free speech communication. It is, however, hard to determine where to install the microphones, because various kinds of acoustic interferences exist inside a car cabin. The noise sources are divided into exterior noise caused by engine, transmission, tire, road surface, and aerodynamic components, and interior noise such as blower noise caused by an in-vehicle air-conditioner. Those noise sources certainly decrease the intelligibility of captured speech in hands-free speech communication.

In this paper, speech intelligibility under the exterior and interior noise conditions is measured using the speech reception threshold (SRT) [5], which corresponds to the sound pressure level of speech at the 50% correct score in an articulation test. In general, SRT rises as noise increases, and comprehensively reflects the relationship between the target speech and acoustical interferences. Speech intelligibility is also modeled in an alternative approach. Multiple regression analyses have been employed to unravel the relationship between speech intelligibility and some subjective indexes related to noisy speech. The seven subjective indexes were carefully prepared as the independent variables in the multiple regression analysis, and participants in a listening test gave the 5-grade mean opinion score on each subjective index. Speech intelligibility is modeled by the linear multiple regression analysis.

[1] mizumach@ecs.kyutech.ac.jp

2. SPEECH RECEPTION THRESHOLD

2.1 Materials

Speech intelligibility can be defined with the utterances of syllables, words, and sentences. In this study, randomly-connected digit sequences are employed as target speech signals, because digit sequences have similar familiarity and non-predictability. Preliminary listening tests were carried out using a wide variety of Japanese digits utterances ranged from single-digit to 7-digits sequences. The single-digit sequences were intensely affected by the local features of noise signals, and the long sequences were suited for testing short-term memory rather than speech intelligibility. It is suggested that 4-digit sequences are the most suitable for investigating speech intelligibility under in-vehicle noisy conditions. Speech materials were prepared from the AURORA-2J database [6].

Acoustical interferences were recorded in real environments using the pre-installed microphone inside the cabin of the hatchback. Exterior roadway noises were recorded on the closed circuit of Mazda Motor Corporation, while the vehicle was cruising at the constant speed of 60 km/h, 80 km/h, 100 km/h, 120 km/h, 150 km/h, and 180 km/h. Interior blower noises were also recorded in the same microphone setting, when the air-conditioner worked with the maximum and minimum airflows. Three different directions of air blow were set to the defroster mode (DEF), the normal ventilation mode (VENT), and the ventilation hitting on the microphone (HIT). In both DEF and VENT modes, the blower noises were almost stationary. On the other hand, spouted air in the HIT mode caused non-stationary bubbling noises. In total, 36 kinds of acoustical interference conditions were prepared with 6 cruising speeds and 6 blower modes, that is, 2 levels of airflow with 3 air blowing directions.

The speech and noise signals were mixed in a computer at the designated signal-to-noise ratio. In each noise condition, five kinds of 4-digits utterances were prepared with different sound pressure levels at the step of 4 dB.

2.2 Procedure

13 Japanese male graduate students and undergraduates with normal hearing participated in a listening test, where the prepared 4-digit sequences were randomly presented through headphones. They were asked to write down the recognized digits in the answer sheet. If a part of the 4-digits sequences was not recognized, the participants partially answered only the recognized digits. Recognition rate is defined as the percentage of the recognized digits in each noise condition. Finally, SRT is obtained form the sigmoid curve fitted to the recognition rates as the function of the sound pressure level of speech.

2.3 Experimental results

Figures 1 and 2 show the recognition rates and the sigmoid functions fitted with non-linear least squares method at the cruising speeds of 60 km/h and 180 km/h, respectively. The recognition rates are averaged over the whole participants. SRT can be obtained from the fitted sigmoid curve as the sound pressure level of which recognition rate is 50 %. It means that a smaller SRT yields high intelligibility. SRTs in all noise conditions are summarized in Fig. 3.

2.4 Discussion

There is no significant difference in the cruising speed in Fig. 3. It could be perceived the louder exterior roadway noise as the cruising speed increased, when the exterior noise was solely presented. The interior blower noises dominate speech intelligibility. Concerning the blower modes, the non-stationary “HIT” mode caused more increase of SRT than the stationary “DEF” and “VENT” modes. It is also found that the resultant SRTs rise more than 20 dB in the “high” levels compared to the “low” levels on the amount of airflow.

For further investigation, SRTs were obtained from the results of another listening tests with single-digit sequences and sentences. The relationship among the SRTs in the 36 noise conditions did not change depending on the length of the target speech, although the SRTs had wide variances in the single-digit condition. In the case of the sentence utterance, SRTs were almost equal to the results with the 4-digits sequences. It suggests that SRT, that is, speech intelligibility, can be efficiently estimated using compact 4-digits sequences instead of uncertain sentence utterances, which involve wide ranges of familiarity of context, speaker individuality, dialect, language, and speaking rate.

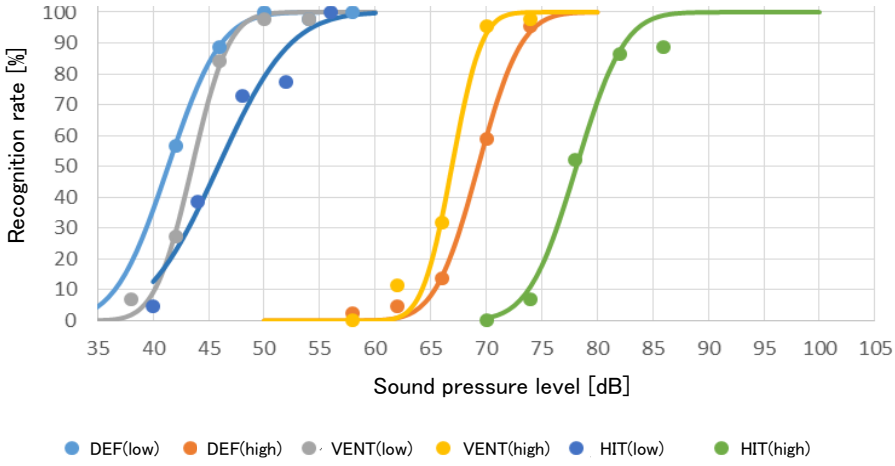

Figure 1 – Recognition rates for Japanese 4-digits sequences while cruising at 60 km/h.

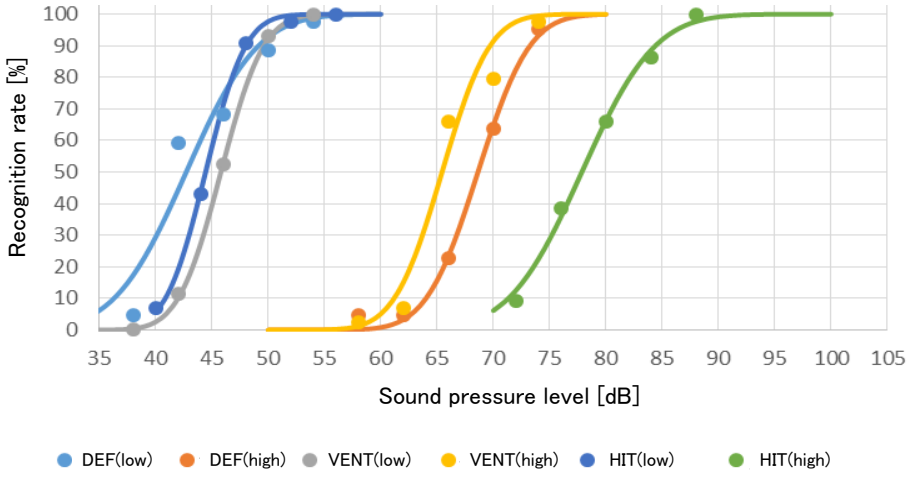

Figure 2 – Recognition rates for Japanese 4-digits sequences while cruising at 180 km/h.

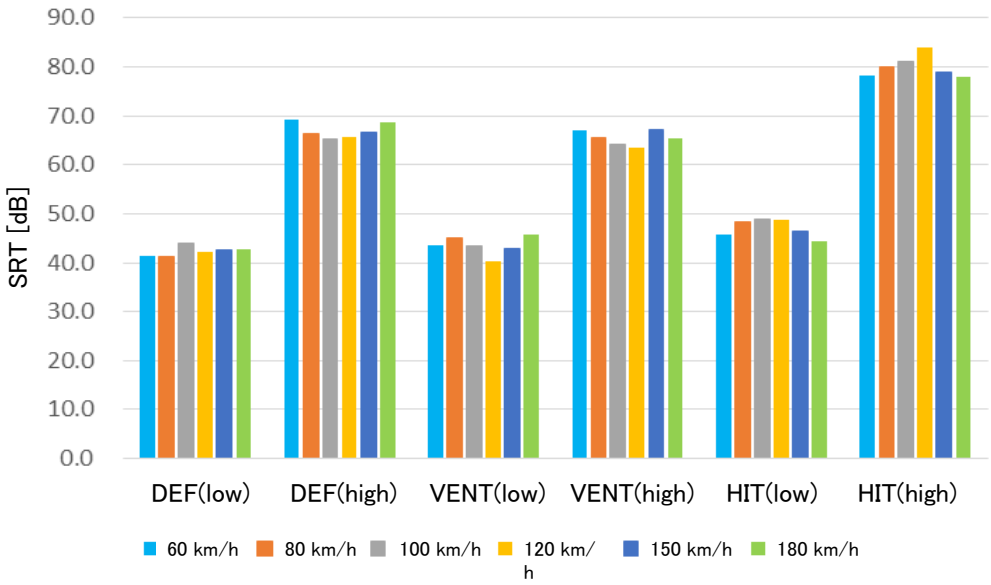

Figure 3 – Speech reception thresholds for Japanese 4-digits sequences.

Table 1 – Dependent and independent variables for multiple regression analysis.

Dependent variable	Y	Speech intelligibility (how easy do you recognize speech?)
Independent variables	X_1	Degree of fluctuation in loudness
	X_2	Degree of thick impression
	X_3	Degree of reverberant impression
	X_4	Loudness of noise
	X_5	Loudness of speech
	X_6	Degree of easiness in dictation
	X_7	Degree of intermittence in speech perception

3. MULTIPLE REGRESSION ANALYSIS

3.1 Materials

Japanese 4-digits sequences were used for the target speech. There was another parameter of speaking rate: fast and slow. Each 4-digits sequence consists of four randomly-selected single-digit utterances, and the prepared 4-digits sequences do not overlap with others.

The exterior roadway noises have less influence on speech intelligibility, and then the exterior noise only at the speed of 60 km/h is considered in this analysis. It is desirable to investigate the effects of the interior blower noises in detail. The blower noises in DEF(low), VENT(low), VENT(high), HIT(low), HIT(mid), and HIT(high) modes, are prepared in three different types of vehicles: hatchback, sedan, and SUV.

The target speech was mixed with the exterior and interior noises in a computer. The sound pressure level of the target speech was set around the noise-dependent SRT. Signal-to-noise ratios range from (SRT-6) dB to (SRT+6) dB at the step of 3 dB in each noise condition.

3.2 Procedure

A listening test was carried out with the cooperation of 10 Japanese male graduate students and undergraduates with normal hearing, where the participants listened to the noisy 4-digit sequences through headphones and gave 5-grade mean opinion score (MOS) on each subjective index. The subjective indexes for multiple regression analysis include the comprehensive intelligibility as the dependent variable and seven independent variables described in Table 1. Linear regression was performed using the averaged 5-grade MOSs for the dependent and seven independent variables.

3.3 Experimental results

The relationship between speech intelligibility (Y) and individual subjective impressions ($X_i, i = 1,2,\cdots,7$) is modeled as follows:

$$Y = -0.07X_1 + 0.12X_2 - 0.03X_3 - 0.09X_4 + 0.43X_5 + 0.53X_6 + 0.11X_7 - 0.04 \qquad (1)$$

where the coefficient of determination is 0.966. Table 2 shows the correlation matrix among the independent variables and *p*-values for the independent variables.

3.4 Discussion

There is high correlation, 0.87, between the independent variables: X_5 and X_6. Either of them should be excluded from the independent variables, and other independent variables with higher *p*-values have been also omitted in turn. Finally, X_2, X_4, and X_6 are used for modeling the speech intelligibility. Those three independent variables have statistical significance at the level of 5 %.

Table 2 – Correlation matrix and P-values for the independent variables.

	Correlation matrix							*p*-value
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	
X_1	1.00	-	-	-	-	-	-	0.501
X_2	0.71	1.00	-	-	-	-	-	0.221
X_3	0.39	0.37	1.00	-	-	-	-	0.888
X_4	0.49	0.44	0.18	1.00	-	-	-	0.003
X_5	0.16	0.30	0.16	0.16	1.00	-	-	0.000
X_6	-0.02	0.20	0.12	-0.20	0.87	1.00	-	0.000
X_7	0.54	0.26	0.12	0.33	-0.55	-0.70	1.00	0.209

In this case, the linear regression model takes the following form:

$$Y = 0.28X_2 - 0.27X_4 + 1.04X_5 - 0.37 \quad (2)$$

where the coefficient of determination is 0.93. It is natural that the loudness of speech (X_5) is the most dominant independent variable and the loudness of noise (X_4) has negative correlation with the speech intelligibility (Y).

4. CONCLUSIONS

Speech intelligibility is considered in the situation of hands-free distant-talking communication inside a car cabin. Speech reception thresholds for the utterances of the 4-digits random sequences point out that speech intelligibility is influenced by the interior blower noise, particularly amount and direction of airflow. As the results of the linear multiple regression analyses with subjective indexes on noisy speech, speech intelligibility can be modeled with three independent variables, that is, thickness of speech, loudness of speech, and loudness of noise.

Future works include investigating the relationship among subjective speech intelligibility and objective acoustic parameters of noisy speech signals. It is also interesting to design a universal intelligibility test signal, which does not depend on speaker, dialect, and language.

REFERENCES

1. French, N. R., & Steinberg, J. C. Factors governing the intelligibility of speech sounds. J. Acoust. Soc. Am., 19(1), 90-119, 1947.
2. ANSI: Methods for the calculation of the articulation index, ANSI S3.5-1969, 1969.
3. ANSI: Methods for calculation of the speech intelligibility index, ANSI S3.5-1997. 1997.
4. Steeneken, H.J.M., Houtgast, T., Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics. Acustica, 46, 60-72, 1973.
5. Plomp, R., & Mimpen, A. M. Speech reception threshold for sentences as a function of age and noise level. J. Acoust. Soc. Am., 66(5), 1333-1342, 1979.
6. Nakamura, S., Takeda, K., Yamamoto, K., Yamada, T., Kuroiwa, S., Kitaoka, N., Nishiura, T., Sasou, A., Mizumachi, M., Miyajima, C., Fujimoto, M., & Endo, T. AURORA-2J: An evaluation framework for Japanese noisy speech recognition. IEICE Trans. Inf.&Syst., 88(3), 535-544, 2005.