Intelligibility Investigation of Single-Channel Noise Reduction Algorithms for Chinese and Japanese

Junfeng LI, Lin YANG and Yonghong YAN
Institute of Acoustics
Chinese Academy of Sciences

Chau Duc THANH and Masato AKAGI
School of Information Science
Japan Advanced Institute of Science and Technology

***Abstract*—A large number of single-channel noise reduction algorithms have been proposed based largely on mathematical principles and evaluated with English speech. Given the different perceptual cues used by native listeners of different languages, it is of great interest to examine whether there are any language effects on speech intelligibility when the same noise reduction algorithm is used to process noisy speech in different languages. In this paper, a comparative evaluation is taken of various single-channel noise reduction algorithms applied to noisy speech for Chinese and Japanese. Clean speech signals (Chinese words and Japanese words) were first corrupted by three types of noise at two signal-to-noise ratios and then processed by five single-channel noise reduction algorithms. The processed signals were finally presented to normal-hearing listeners for recognition. Intelligibility evaluations showed that the majority of noise-reduction algorithms did not improve speech intelligibility and that significant differences in performance of noise reduction algorithms were observed across the two languages.**

I. Introduction

In real environments, speech signals are often corrupted by various kinds of noises. To eliminate the effects of noises on speech, many single-channel noise reduction algorithms have been proposed [1]. Typical algorithms are subspace, statistical-model based, spectral subtractive and Wiener-type algorithms [1]. Although these algorithms have shown great effectiveness in suppressing background noise and improving speech quality, no significant benefit, if any, in improving speech intelligibility was observed relative to unprocessed noisy signals using English speech in noisy conditions [2].

The field of linguistics suggests that different languages are generally characterized by diverse specific features at the acoustic and phonetic levels, due to their distinctive production manner, perceptual mechanism and syntax structures [3]. Compared with English, for example, Chinese and Japanese contain much fewer vowels, which results in more severe phoneme and syllable confusions for Chinese and Japanese in noise. Furthermore, the F0 information in English is used primarily to emphasize or express emotion and contribute little to speech intelligibility, at least in quiet. In contrast, the tone information (as carried in the F0 contour) in Chinese and the accent information in Japanese, are used to distinguish word meaning and thus contribute a great deal to Chinese and Japanese speech intelligibility [3].

Part of this work was done when Dr. Junfeng Li was an Assistant Professor at Japan Advanced Institute of Science and Technology.

The differences among languages have been extensively studied in the context of speech recognition in noise [4], [5], [6], [7]. When demonstrating the primary role of temporal envelope in speech recognition, speech recognition of English was investigated by preserving temporal envelope cues but removing the spectral detail within each frequency band [4]. Following the same methodology, Fu *et al.* investigated the role of temporal envelope cues in Chinese and showed that the high level of tone recognition (about 80% correct) in Chinese yielded a significant difference in speech recognition between Chinese and English when no spectral information was available [5]. When demonstrating the effectiveness of the rapid speech transmission index (RASTI), Houtgast *et al.* tested the RASTI across ten Western languages and showed that language-specific effects could be a factor resulting in disparity among diverse tests [6]. Moreover, Kang performed a series of intelligibility tests in two different enclosures and reported that speech intelligibility of English is considerably better than intelligibility of Mandarin speech in noisy backgrounds[7].

Due to the great variation in speech recognition among different languages, it is important to assess the performance in speech intelligibility of noise reduction algorithms for different languages under various noisy conditions. Accordingly, in this paper, phonetically-balanced Chinese and Japanese words corrupted by three different types of noise were first processed by five single-channel noise reduction algorithms and then presented to native Chinese and Japanese listeners respectively for word identification. The contributions of this research are: (1) These evaluations will help us understand which algorithm(s) preserves or enhance speech intelligibility relative to that of unprocessed signals for Chinese and Japanese. (2) More importantly, this study will examine any performance differences in speech intelligibility of existing noise reduction algorithms when applied and used in different languages.

II. Intelligibility investigation of noise-reduction algorithms for Chinese

A. Subjects

Ten native Mandarin listeners (five females and five males) with normal hearing, aged from 23–31 years old, participated in our experiment. They were paid for their participation.

B. Materials

In the intelligibility evaluations of single-channel noise reduction algorithms for Mandarin, the syllable tables for intelligibility test reported by Ma *et al.* were adopted as our materials [8]. This set of test materials consists of 10 syllable tables, each of which contains 75 phonetically balanced (PB) Mandarin syllables with CV (Consonant-Vowel) structure. In each table every three syllables are combined randomly, generating nonsense sentences with the format "The ith sentence is *word1*, *word2*, *word3*". Consequently, every table can produce enough lists with 25 nonsense sentences to conduct general tests. The sentence lists were recorded in a sound-proof booth at a sampling rate of 16 kHz and stored in a 16-bit format, and then down-sampled to 8 kHz before being presented to the listeners.

C. Signal processing

The clean and noise signals were processed by the IRS filter to simulate the receiving frequency characteristics of telephone network, which is similar to the study in [2]. Then the noise signals were added to the clean speech at 0 dB and 5 dB SNRs. Three types of background noises, white noise, babble noise and car noise, were used in our evaluations.

The noisy signals were enhanced by five representative single-channel noise reduction algorithms, including KLT, logMMSE, logMMSE-SPU, MB and Wiener-as [1], [2]. These algorithms cover the state-of-the-art four major classes of noise reduction: subspace (KLT), statistical-model based (logMMSE and logMMSE-SPU), spectral subtractive (MB) and Wiener-type algorithm (Wiener-as). Matlab implementation of these algorithms are available in [1].

D. Procedure

The noisy and enhanced signals were presented to the subjects at a comfortable level through TDH-39 headphone and Madsen Iteral II audio meter in a sound-proof booth. All subjects went through a training procedure to become familiar with the testing environment. In the formal tests, there were 36 listening conditions, including 3 types of background noises (white, babble and car noises) $\times$ 2 SNR levels (0 dB and 5 dB) $\times$ 6 enhancement types (noisy reference and five noise reduction algorithms). Every subject would listen to 900 (25 $\times$ 36) nonsense short sentences. All listening conditions were divided into three parts according to the background noise type, and in each listening part the sentences were randomly presented to the subjects. Listeners were asked to write down the keywords that they heard in every sentence.

E. Results

Fig. 1 shows the word recognition scores averaged across ten subjects for five enhancement algorithms under three background noises at two SNR levels. The error bars represent the standard errors of the mean. As shown in Fig. 1, the word recognition scores at 0 dB SNR levels were much lower than those at 5 dB SNR levels for all algorithms in the tested noise conditions. For each noise reduction algorithm, the recognition rates in babble noise conditions were much lower than those in white and car noise conditions. One main reason for this was that it was much more difficult to estimate babble noise spectrum due to its non-stationarity. Of all the algorithms, the logMMSE-SPU algorithm provided the lowest intelligibility scores in all noise conditions, which could be attributed to the severe speech distortion introduced by its noise reduction processing. In most conditions, speech intelligibilities of the processed signals by the tested algorithms (except for the Wiener-as algorithm) were much lower than that of the unprocessed noisy signals. The Wiener-as algorithm provided almost equivalent speech intelligibility, or a small degree of improvement, compared with the noisy signals in most listening conditions.

III. Intelligibility investigation of noise-reduction algorithms for Japanese

A. Subjects

Thirty normal-hearing listeners (twenty-seven males and three females) participated in this experiment. All subjects were native listeners of Japanese, and were paid for their participation. The subjects' age ranged from 23 to 36 years old.

B. Stimuli and signal processing

In the intelligibility evaluations for Japanese, the words taken from the familiarity-controlled word lists 2003 (FW03) were used as speech material, which consisted of 80 lists with 50 phonetically-balanced words per list [9]. Since word familiarity has a strong effect on word recognition, all word lists in FW03 were divided into four sets in four word-familiarity ranks. All words were recorded in a soundproof room at a 48 kHz sampling rate. In the present investigation, only the word lists with the lowest familiarity uttered by one female were used to be comparable with the nonsense words used in the evaluations for Chinese.

The noise signals in the intelligibility evaluations for Japanese were the same as those used in Chinese intelligibility evaluations. As in the investigation for Chinese, speech and noise stimuli were first downsampled to 8 kHz and filtered by IRS filters, and then mixed to generate the noise-corrupted signal at SNRs of 0 and 5 dB. The noise-corrupted signals were finally processed by five single-channel noise reduction algorithms as used in the intelligibility evaluations for Chinese.

C. Procedure

Thirty listeners were grouped into three panels (one panel per type of noise), with each panel consisting of ten listeners (one female and nine males). Each panel of listeners listened to words corrupted by a different type of noise, which ensures that each subject listened to each sentence once. The noisy and processed signals were presented to the subjects at a comfortable listening level through HDA-200 headphone in a sound-proof booth. Prior to the test, each subject went through a training procedure to become familiar with the testing procedure. In the tests, each subject participated in a total of 12

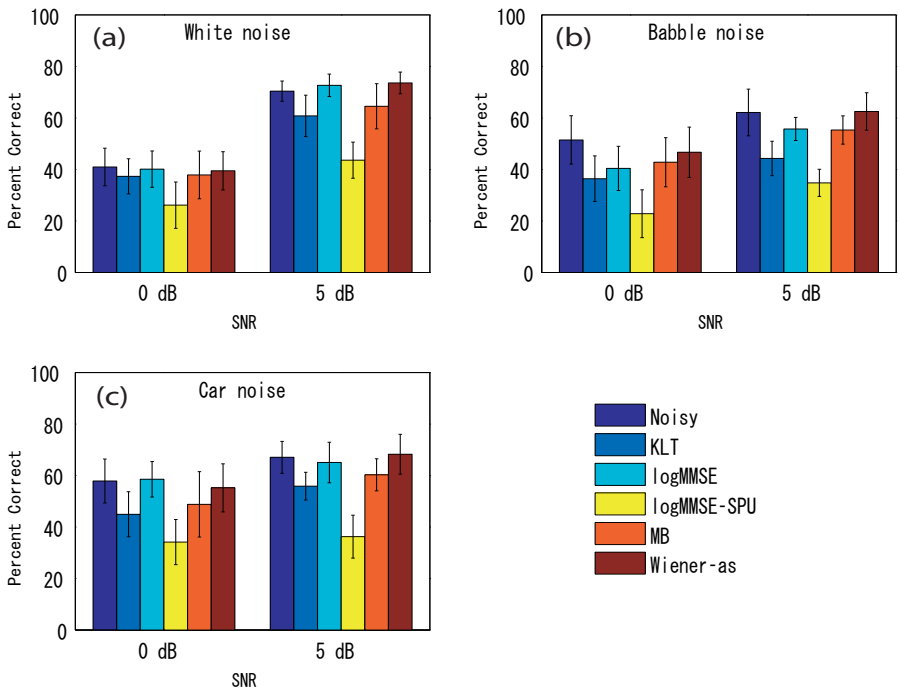

Fig. 1. *Mean word recognition scores of five enhancement algorithms for Mandarin Chinese under three types of background noise at two SNRs.*

listening conditions [=2 SNR levels $\times$ 1 types of background noise $\times$ 6 algorithms (5 noise-reduction algorithms + 1 unprocessed references)]. One word list of 50 words were used for each condition. Each subject listened to 600 low-familiarity words (=50 sentences $\times$ 12 conditions) in the listening test. For each panel, the presentation order of the stimuli and listening conditions were randomized across each subject. Subjects were asked to write down the words they heard.

D. Results

The mean performance in terms of percentage of Japanese words identified correctly across subjects for six processing conditions under three noise scenarios at two SNRs are plotted in Fig. 2, in which the error bars represent the standard errors of the mean. Similar to the Mandarin intelligibility scores, the word recognition scores of Japanese at 0 dB were also greatly lower than those at 5 dB for all processing conditions in the tested noise environments, as indicated in Fig. 2. Again, the logMMSE-SPU algorithm yielded the lowest word recognition scores among the tested algorithms under all noise conditions. In babble noise conditions, all tested noise-reduction algorithms decreased speech recognition rate compared with the unprocessed noisy signal. In white and car noise conditions, the logMMSE algorithm yielded the same recognition rates as those of the noisy signals, and the Wiener-as algorithm provided a small degree of improvements compared with the noisy signals. Other algorithms showed a significant decrease in word recognition rate relative to the noisy signals.

IV. Intelligibility Comparisons between Chinese and Japanese

One of the important purposes of this study is to compare the different effects of the noise reduction algorithms on speech intelligibility in different languages. Due to the fundamental differences among languages, it is unreasonable to directly compare the absolute word identification scores among different languages [7]. As one alternative approach, the comparison can be made indirectly by comparing the difference of word identification scores from one condition to another [7]. In this research, the relative speech intelligibility that was derived by subtracting the word identification score of the processed speech by the noise reduction algorithms from that of the unprocessed noise-corrupted speech was used for speech intelligibility comparison.

The relative word recognition scores across subjects of all tested noise reduction algorithms in two noise conditions (babble and car noises) at two SNRs (0 and 5 dB) for two languages (Chinese and Japanese) are shown in Fig. 3. As can be seen from Fig. 3, there was a large variability in performance with the noise reduction algorithms even when tested in the same noise conditions. In the 0 dB babble noise, for example, the Wiener-as algorithm showed a positive (though a small degree) improvement in word identification relative to the noise-corrupted speech for Japanese, and a negative improvement for Chinese. Though the logMMSE-SPU algorithm was proven very effective in improving speech quality of English [2], it yielded the worst ability in speech intelligibility enhancement for both Chinese and Japanese.

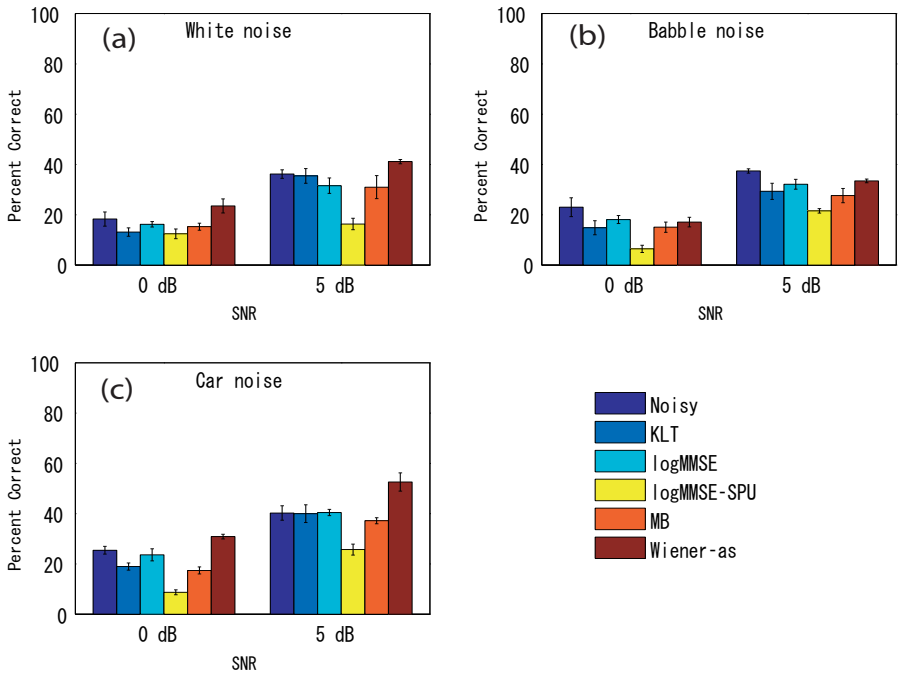

Fig. 2. *Mean word recognition scores of five enhancement algorithms for Japanese under three types of background noise at two SNRs.*

Moreover, although most of tested algorithms (except for the Wiener-as algorithm) yielded the decrements in the relative word recognition rate, the degree of degradation was largely varying between Japanese and Chinese.

To further understand the performance difference of each noise reduction algorithm between Chinese and Japanese in different conditions, Multiple paired comparisons with Ryan's method were conducted between the relative recognition scores for different languages. Difference between the relative word recognition scores were regarded as significant if the significance level $p < 0.05$. The analysis results are listed in Table I. The performance in relative word identification scores of the tested noise-reduction algorithms exhibited significant difference in relative word identification scores between languages in a majority of conditions. The KLT algorithm showed significant difference in relative word recognition scores between two languages in all high SNR conditions and 0 dB car noise condition. The significant performance differences of the logMMSE algorithm was only observed under 5 dB white noise condition. While the logMMSE-SPU algorithm demonstrated significant difference for two languages in all tested conditions. The additional significant differences between two languages were also found for the Wiener-as algorithm in 0 dB white noise and car noise conditions. While the MB algorithm showed no significant performance differences in all tested conditions.

TABLE I
MULTIPLE PAIRED COMPARISON RESULTS OF LANGUAGE FACTORS (CHINESE AND JAPANESE). THE ASTERISKS "*" INDICATED THE SIGNIFICANT PERFORMANCE DIFFERENCES IN THE WORD RELATIVE INTELLIGIBILITY SCORES UNDER THE CORRESPONDING CONDITIONS.

	White		Babble		Car	
Algorithm	0 dB	5 dB	0 dB	5 dB	0 dB	5 dB
KLT		*		*	*	*
logMMSE		*				
logMMSE-SPU	*	*	*	*	*	*
MB						
Wiener-as	*				*	*

V. CONCLUSIONS

In this paper, two experiments were carried out to examine the intelligibility of signals processed by five noise reduction algorithms for Chinese and Japanese under three types of noise at two SNRs. Based on the investigation results in word identification scores, the following observations were obtained:

1) In severe noise conditions, most single-channel noise-reduction algorithms were unable to recover the temporal and spectral characteristics of consonants, resulting in low word recognition scores. .
2) Considering all the conditions examined, the Wiener-as algorithm performed the best in that it either maintained or improved word recognition scores relative to unprocessed noisy speech for Chinese and Japanese. The logMMSE algorithm ranked second in speech in-

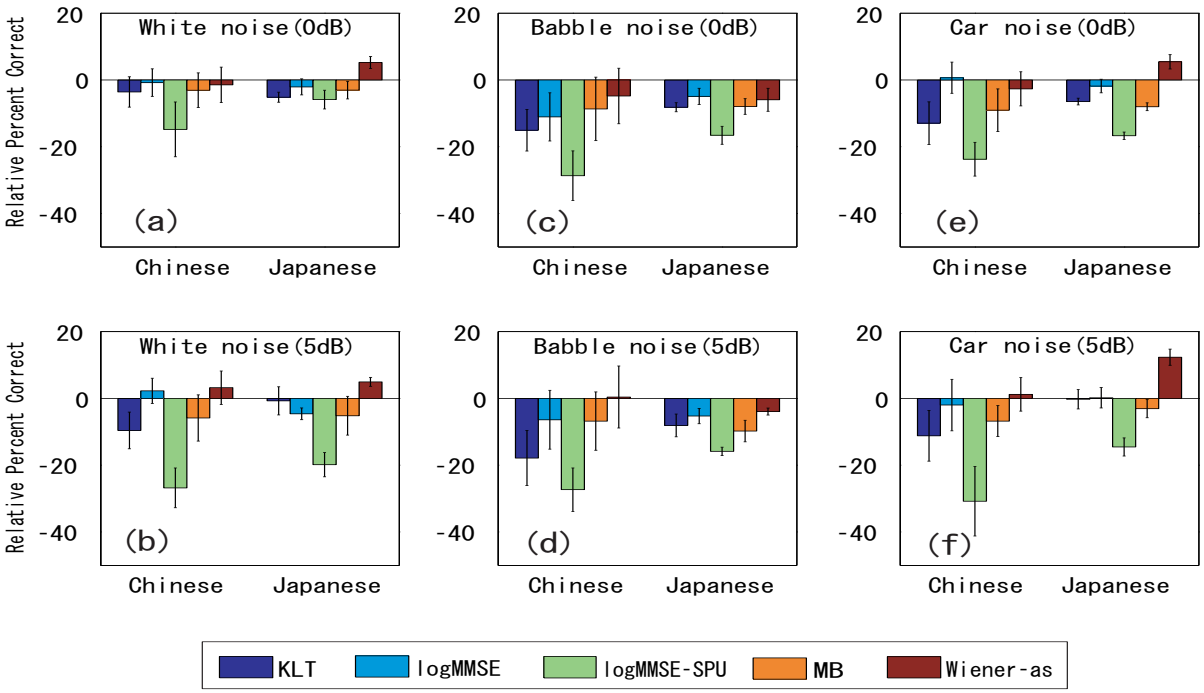

Fig. 3. *Intelligibility comparison of single-channel noise-reduction algorithms for Chinese and Japanese under three (white, babble and car) noise conditions.*

telligibility followed by the KLT and MB algorithms. The logMMSE-SPU algorithm yielded the worst performances in all tested conditions. This ranking of different noise-reduction algorithms is consistent for Chinese and Japanese.

3) The differences in relative word recognition score for different languages demonstrated that most noise reduction algorithms were significantly affected by the language in most conditions. In the extremely difficult conditions (e.g., the 0 dB babble noise), only the logMMSE-SPU algorithm showed significant difference between two languages. One possible reason for this is that the important perceptual speech cues (e.g., formant information, F0 information) for word recognition were too difficult to be extracted from the processed signal by most noise reduction algorithms.
4) Significant differences in relative word recognition score between Chinese and Japanese were found for the KLT, logMMSE-SPU and Wiener-as algorithms in most conditions, and for the logMMSE algorithm in only 5 dB white noise condition. No significant difference was found for the MB algorithm in all tested conditions.

VI. Acknowledgements

Research is partially supported by National Science & Technology Pillar Program (2008BAI50B00), National Natural Science Foundation of China (10874203, 60875014, 60535030, 11074275), and the China-Japan (NSFC-JSPS) Bilateral Joint Projects.

References

[1] P.C. Loizou, "Speech enhancement:Theory and practice", CRC Press, 2007.
[2] Y. Hu and P.C. Loizou, "A comparative intelligibility study of single-microphone noise reduction algorithms", J. Acoust. Soc. Am., 122(3):1777-1786, 2007.
[3] R.L. Trask, "Key concepts in Language and Linguistics," Routledge, 1998.
[4] R.V. Shannon, F.G. Zeng, V. Wygonski and M. Ekelid, "Speech recognition with primarily temporal cues," Science, 270, pp. 303-304, 1995.
[5] Q.J. Fu, F.G. Zeng, R.V. Shannon and S.D. Soli, "Importance of tonal envelope cues in Chinese speech recongition," J. Acoust. Soc. Am., 104(1):505-510, 1998.
[6] T. Houtgast, H. Steeneken, "A multi-language evaluation of the rasti method for estimating speech intelligibility in auditoria," Ascoustica, 54, pp. 185-199, 1984.
[7] J. Kang, "Comparison of speech intelligibility between English and Chinese," J. Acoust. Soc. Am., 103(2):1213-1216, 1998.
[8] D. Ma and H. Shen, "Acoustic Manual", Chinese Science Publisher, 527-537, 2004.
[9] S. Amano, S. Sakamoto, T. Kondo, Y. Suzuki, "Development of familiarity-controlled word lists 2003 (FW03) to assess spoken-word intelligibility in Japanese," Speech Communication, 51, pp. 76-82, 2009.