A Sound Masking Signal Generation Method Based on Chinese Pronunciation Characteristics

Xiaofeng MA[1, a], Peng ZHANG[1, b], Qiuyun HAO[1, c] ,Xiaoxia CHEN[2], Yanhong FAN[1], Jingsai JIANG [1], Ye LI[1]

[1]Shandong Provincial Key Laboratory of Computer Networks, Shandong Computer Science Center (National Supercomputer Center in Jinan), Jinan, 250014, China

[2]Northwestern Polytechnical University, Xian, 710072, China

[a]email: maxf@sdas.org, [b] email:zhangp@sdas.org, [c]email: haoqy@sdas.org

Keywords: Sound masking; Speech privacy; Chinese pronunciation characteristics

Abstract. Sound masking is a technique to protect the speech privacy of confidential talks in a room. Most of the existing masking signals sometimes become noise sources, and result in a negative impact on the people in the room. This paper proposes a new method for generating masking signals. The synthesis of the new masking signal is based on Chinese pronunciation characteristics and the statistical properties. Since its statistical properties are similar to those of human's voice, the proposed method is more difficult to be deciphered, and can improve the masking effect and reduce the impact on people's psychology and physiology.

Introduction

Conference rooms often appear to provide speech confidentiality but actually may not [1]-[3]. Speech confidentiality is important to the information security of nation, business and civilian. As the most basic information, voice is the focus that needs to be protected. Sound masking can be used to compensate for the acoustic weakness of secure rooms. Currently, the common masking signal mainly includes white noise, pink noise, HVAC noise, etc [4]-[6]. White noise and pink noise usually have relatively stable statistical characteristics but less masking efficiency, while HVAC noise is discontinuous, unstable and uneven, or the sound level is too high. Sometimes the masking sound becomes a noise source, and results in a negative impact on people's psychology and physiology [7]-[9]. Considering this problem, we propose a new method for generating the masking signal. The synthesis of the new masking signal is based on the Chinese pronunciation characteristics and the statistical properties, including words, phrases and sentences. As its statistical properties are similar to those of the normal voice, it is difficult to be deciphered, and thus improves the masking effect and reduces the impact on people's psychology and physiology.

Proposed Sound Masking Signal Generation Method

1) Generation of Random Text

Chinese speech contains several elements: paragraphs, sentences, segments, phrases, and words. Generation of the random text involves several statistics derived from the modern Chinese universal balancing library. The library was set up by Chinese National Committee of Language, and contains about 100 million words with a large span of time, an extensive distribution, and more balanced proportion. It could be a panorama of modern Chinese language.

a) Sentence probability table. Sentence probability table can be obtained according to the statistical analysis of the number of sentences in each paragraph, and is denoted by $\left[J_1, J_2, J_3 \cdots, J_m\right]$, where $J_i\ (1 \le i \le m)$ represents the percentage of paragraphs containing i sentences;

b) Segment probability table. Segment probability table can be obtained according to the statistical analysis of the number of segment s in each sentence, and is denoted by

$[D_1, D_2, D_3 \cdots, D_l]$, where D_i $(1 \le i \le l)$ represents the percentage of sentences containing i segments;

c) Phrase probability table. Phrase probability table can be obtained according to the statistical analysis of the number of Phrases in each segment, and is denoted by $[C_1, C_2, C_3 \cdots, C_q]$, where $C_i (1 \le i \le q)$ represents the percentage of segments containing i phrases ;

d) Word probability table. Word probability table can be obtained according to the statistical analysis of the number of words in each phrase, and is denoted by $[Z_1, Z_2, Z_3 \cdots, Z_p]$, where $Z_i (1 \le i \le p)$ represents the percentage of phrases containing i words ;

e) Syllable probability table. All syllables are sorted in alphabetical order, denote as $[H_1, H_2, H_3 \cdots, H_k]$, and syllable probability table can be obtained according to the probability of each syllable appears in everyday speech, denote as $[h_1, h_2, h_3 \cdots, h_k]$, where h_i $(1 \le i \le k)$ represents the probability of syllable H_i appears in everyday speech.

f) According to the above probability tables, follow the steps below to generate random text message.

f-1) Determine the number of sentences of each paragraph. In the interval $\left[0, \sum_{i=1}^{m} J_i\right]$, a random number r_1 is generated. If $r_1 \in \left[\sum_{i=0}^{n1-1} J_i, \sum_{i=0}^{n1} J_i\right]$ $(1 \le n1 \le m,\ J_0 \neq 0)$, the number of the sentences contained in the paragraph is set equal to $n1$. For example, if $r_1 \in [0, J_1]$, the paragraph contains one sentence; if $r_1 \in [J_1, J_1 + J_2]$, the paragraph contains two sentences, and so on.

f-2) Determine the number of segments of each sentence. In the interval $\left[0, \sum_{i=1}^{l} D_i\right]$, a random number r_2 is generated. If $r_2 \in \left[\sum_{i=0}^{n2-1} D_i, \sum_{i=0}^{n2} D_i\right]$ $(1 \le n2 \le l,\ D_0 \neq 0)$, the number of the segments contained in the sentence is set equal to $n2$. For example, if $r_2 \in [0, D_1]$, the sentence contains one segment; if $r_2 \in [D_1, D_1 + D_2]$, the sentence contains two segments, and so on.

f-3) Determine the number of phrases of each segment. In the interval $\left[0, \sum_{i=1}^{q} C_i\right]$, a random number r_3 is generated. If $r_3 \in \left[\sum_{i=0}^{n3-1} C_i, \sum_{i=0}^{n3} C_i\right]$ $(1 \le n3 \le q,\ C_0 \neq 0)$, the number of the phrases contained in the segment is set equal to $n3$. For example, if $r_3 \in [0, C_1]$, the segment contains one phrase; if $r_3 \in [C_1, C_1 + C_2]$, the segment contains two phrases, and so on.

f-4) Determine the number of words of each phrase. In the interval $\left[0, \sum_{i=1}^{p} Z_i\right]$, a random number r_4 is generated. If $r_4 \in \left[\sum_{i=0}^{n4-1} Z_i, \sum_{i=0}^{n4} Z_i\right]$ $(1 \le n4 \le p,\ Z_0 \neq 0)$, the number of the words contained in the phrase is set equal to $n4$. For example, if $r_4 \in [0, Z_1]$, the phrase contains one word; if $r_4 \in [Z_1, Z_1 + Z_2]$, the phrase contains two words , and so on. Each word corresponds to a syllable, so the number of syllables is equal to the number of words.

f-5) Determine the syllables. In the interval $\left[0, \sum_{i=1}^{k} h_i\right]$, a random number r_5 is generated. If $r_5 \in \left[\sum_{i=0}^{n5-1} h_i, \sum_{i=0}^{n5} h_i\right]$ ($1 \le n5 \le k$, $h_0 \neq 0$), the syllable corresponds to the word is H_{n5}. For example, if $r_5 \in [0, h_1]$, the corresponding syllable is H_1; if $r_5 \in [h_1, h_1 + h_2]$, the corresponding syllable is H_2, and so on

Follow steps f-1) to f-5), a paragraph of random text is generated, repeat the procedure unit all the paragraphs are completed.

2) Synthesis of Masking Signal

Speech synthesis is based on a professional speech database,covering most of the commonly-used Chinese syllables. In the speech database, the syllables are sorted in alphabetical order and named corresponding to their pronunciation and accent, such as “a1.wav” and “ba1.wav”. Match the above random text “text.txt” with the speech database, for example, if the first syllable of the random text is “bai3”, then ”bai3.wav” will be picked out from the speech database. Repeat this procedure until the random text is completed.

Silent periods are added between paragraphs, sentences, segments to obtain more smooth and natural speech. Carriage return and line feed are assigned as the notation of paragraph; period, question mark and exclamation mark are used to denote sentences; and colon, comma and semicolon are placed between segments. Pre-recorded silent segment is added into the speech database and named differently from all other syllables, such as “jyin.wav”. When reading the random text, if the above end symbols are encountered, pick out the silent segment to achieve the specified speech pauses.

Experiments

The experimental materials are selected from the mandarin sentence materials for Chinese speech audiometry, including four groups of phrases and eleven sentences[10]. Synthesize the experimental signal under white noise and the proposed masking noise. The signal-to-noise ratios (SNR) are set equal to -16dB, -12dB, -10dB, -8dB, -4dB, 0dB and 4dB. Experiments are carried out in a professional studio. The experimental signals are played through Cool Edit Pro 2.0. In subjective tests, 30 listeners are employed, all from colleges and without hearing impairment.

The masking performance is measured in speech intelligibility, which can be graded as 3 levels.

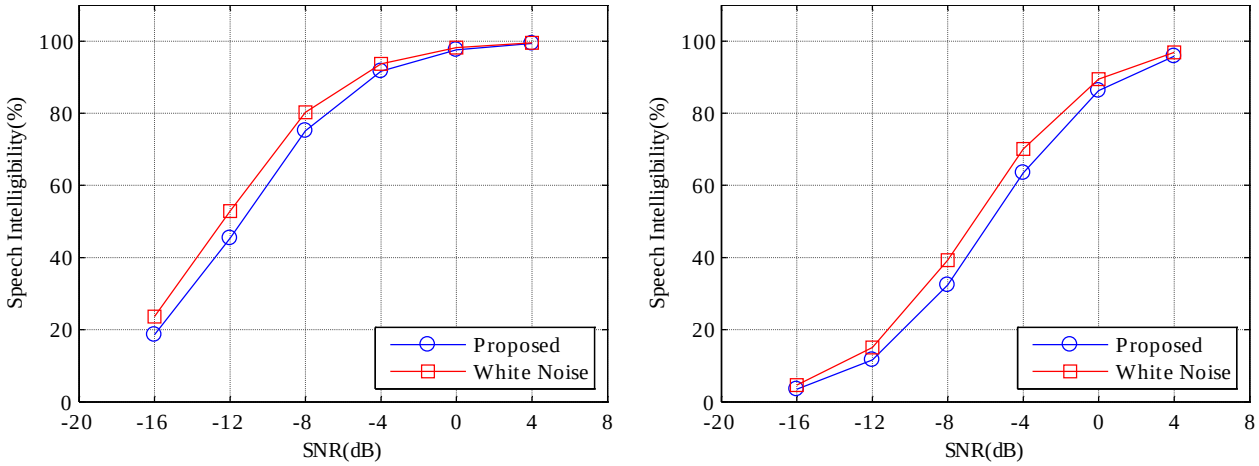
Score 1: speech signal is completely inaudible.

Score 2: the contents of the speech can be heard but can not be understood.

Score 3: the speech can be completely understood.

Lower score means better masking effect and better speech privacy.

Experimental results of each group of phrases data are figured out, and the geometric mean for each group is calculated. Then the correlation coefficient is calculated, the data deviation from the average value of the correlation coefficient is removed, and the geometric mean is recalculated. Sentence experimental signal is processed by the same manner, and all results are converted to speech intelligibility percentile, as shown in Fig. 1.



(a) phrase experiments (b) sentence experiments

Fig1. Speech intelligibility versus SNR for white noise and the proposed masking signal

The results indicate that, compared to the phrase experimental signal, the sentence experimental signal has better the masking effects under the same conditions. In either case, the proposed masking signal performs better than traditional white noise.

Conclusion

In this paper, a new method for generating sound masking signal is proposed. The synthesis of the masking signal is based on the Chinese pronunciation characteristics and the statistical properties, including words, phrases and sentences. Due to its statistical properties which are similar to the normal voice, the proposed method is more difficult to be deciphered, Experimental results show that compared with white noise, the proposed masking signal can improve the masking effect and reduce its disturbance.

Acknowledgement

This work was supported by the International Science & Technology Cooperation Program of China (Grant No. 2012DFR10500), by Shandong Provincial Young and Middle-Aged Scientists Research Awards Fund (Grant No. BS2014DX019), and by the Youth Science Funds of Shandong Academy of Sciences (Grant No. 2014QN009).

References

[1] W.J. Cavanaugh, W.R.Farrell, P.W.Hirtle, et al. Speech privacy in buildings[J]. The Journal of the Acoustical Society of America, 1962, 34(4): 475-492.

[2] IEC268-16.Sound system equipment-Part16: objective rating of speech intelligibility by speech transmission.index, 1997, 12.

[3] Xing Xiaojuan, Jiao Fenglei， Kang Jian, Jin Hong. The Research Progress of Speech Privacy in Open Plan Offices. Noise and Vibration Control.2009, S2:358-362.

[4] Yang Congjing, Liu Mingzhu, Wang Manyuan,Yu Xiao yang.The Research and Development of Masking Sound System. Journal HARBIN UNIV.SCI.＆TECH.2001, 6(4):18-21.

[5] T. Tamesue, S. Yamaguchi, T.Saeki. Study on achieving speech privacy using masking noise[J]. Journal of sound and vibration, 2006, 297(3): 1088-1096.

[6] T.Komiyan.An efficient speech privacy system using speaker-dependent babble noise as masker[C].Internoise2011, Osaka, 2011

[7] V.Hongisto, A.Haapakangas. Effect of sound masking on workers in an open office[C]. Proceedings of Acoustics. 2008, 8: 537-542.

[8] T.Fujii,S.Yamaguchi,T.Saeki. Effects of meaningful or meaningless external noise on participants during simple mental tasks[J]. The Japanese Journal of Ergonomics, 2002, 38(1): 63-8.

[9] T.Saeki,T. Tamesue,S. Yamaguchi, K.Sunada. Selection of meaningless steady noise for masking of speech[J].Applied Acoustics, 2004, 65(29):203-210

[10] Zhang Xiaojie,Cen Wenjuan,Mao Dongxing. Study of Noise Masking Property for Improving Speech Privacy. Noise and Vibration Control. 2012, S1:119-122.