

Title	変調知覚メカニズムに着目した騒音低減法の検討
Author(s)	磯山, 拓都
Citation	
Issue Date	2018-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/15461
Rights	
Description	Supervisor:鵜木 祐史, 先端科学技術研究科, 修士 (情報科学)



Study on noise suppression method based on modulation perception mechanism

Takuto Isoyama (1610008)

Graduate School of Advanced Science and Technology, JAIST,
isoyama-t@jaist.ac.jp

Extended Abstract

Humans perceive various types of sounds at various sound-pressure levels in our daily life. For example, speech and music are perceived as desired sound, and background stationary and non-stationary noise as undesired sound. Heavy noise not only dramatically reduces intelligibility of speech but also induces hearing loss and hearing fatigue in case of long-time exposure. Therefore, noise suppression is important for enhancing speech intelligibility as well as protecting hearing ability.

There are many kinds of noise suppression methods. Classical and popular method for suppressing noise is Boll's spectrum subtraction method. It can successfully suppress stationary components of background noise by subtracting the averaged amplitude spectrum from noisy speech. However, this method cannot sufficiently suppress non-stationary noise such as impulsive noise and intermittent noise.

Several methods have been proposed for suppressing non-stationary noise. One of them can sufficiently suppress impulsive noise by using zero-phase signal. However, the drawback of that method is that the impulsive components of the unvoiced signals (consonants) are also being removed. Non-negative spectral decomposition was proposed to suppress both stationary and non-stationary noise. However, this method trains noise properties using a preliminary learning technique, so noise reduction is limited to the training data. It is difficult to reduce both stationary and non-stationary noise simultaneously without prior knowledge of the noise types and preliminary learning.

From knowledge of human auditory perception, temporal modulation can be regarded as an important part of speech perception as well as of sound quality assessment. Therefore, the author's motivation is to mimic noise reduction based on auditory modulation perception. This paper proposed a method for suppressing both stationary and non-stationary noise based on modulation perception mechanism.

Gammatone filterbank was used for modulation spectral analysis of stationary, impulsive and intermittent noise. As a result, It was reconfirmed that the modulation spectrogram of speech stimuli have a unique peak around modulation frequency of 4 Hz. It was found that the modulation spectrogram of stationary noise such as white noise, pink noise, and babble noise appear in the lower modulation frequencies. It was also found that modulation spectrogram of machine

gun noise as intermittent noise appears as harmonics. From the analyses of the datasets, it was found that the fundamental modulation frequency of the machine gun noise was 8 Hz, while the modulation spectrogram of the impulsive noise appears as a flat shape with the dynamic range of 5 dB in all modulation frequencies. These values of the f_m and dynamic range depend on the datasets, so they should be automatically determined by the auto-correlation technique. These features were then used to suppress the stationary and non-stationary noise components from the observed signals.

Using the Therefore processing, the direct-current components of the spectrogram in the stationary noise, harmonicity of the spectrogram in the intermittent noise, and higher modulation-frequency components of the spectrogram in the impulsive noise were removed. (1) It is found that the modulation spectrogram of stationary noise appears in the lower modulation frequencies. Thus, to obtain the power envelope with removed stationary noise, the DC component of the modulation spectrogram was cancelled out by using the following processing. (2) It is found that the modulation spectrogram of intermittent noise appears as harmonics with the fundamental modulation frequency of 8 Hz. Thus, to remove the intermittent noise component, these harmonics of the modulation spectrogram are canceled out by the following finite impulse response (FIR) band-stop filtering. (3) It is found that the modulation spectrogram of impulsive noise appears in the entire modulation frequency domain as a flat shape. Thus, to remove the impulsive noise component, the modulation spectrogram shape is attenuated by using the low-pass filter.

Five types of objective measures were used to evaluate the proposed method. The first two measures evaluated the efficiency of the proposed method in suppressing the noise components. First, one of them was used to evaluate the noise suppression level. Second, another measure was used to evaluate the relative suppression level. The last three measures were psychoacoustical sound-quality indices, loudness, sharpness, and roughness. These measures were used to objectively evaluate sound-quality after noise-suppression. Loudness indicates the attribute of a sound that determines the magnitude of the auditory sensation produced. Sharpness and roughness indicate complex effects that quantify the subjective perception of rapid and sharp sound. Thus, heavy noise, in general, induces increasing loudness, increasing sharpness, and increasing roughness.

It was found that the proposed method can suppress stationary noise by 8 dB, intermittent noise by 6 dB, and impulsive noise by 8 dB terms of the suppression level.

It was also found that the proposed method can sufficiently suppress the noise effects from noisy speech by 8 dB as SNRs from -20 to -60 dB in terms of relative suppression level.

It was found that when the sound pressure level of noise is 100 dB, reduced loudness of stationary noise is 50 sone, while that of intermittent noise and impulsive noise is 20 sone. In addition, it was found that reduced loudness increases as the sound pressure level of noise increases.

It was found that when the sound pressure level of noise is 100 dB, reduced sharpness of stationary noise is 0.1 acum while that of intermittent noise and impulsive noise is 0 acum.

It was found that when the sound pressure level of noise is 100 dB, reduced roughness of stationary noise is 0.05 asper, that of intermittent noise is 0.73 asper, and that of impulsive noise is 0.25 asper.

All of the results confirmed that the proposed method can perceptually reduce the noise effects for speech enhancement, even if the sound pressure level of noise is high. From the evaluation results above, it is indicated that the proposed method can sufficiently suppress stationary and non-stationary noise. Moreover, it can also reduce the perceptual effects due to noise exposure.