Title	音声明瞭度向上のための残響音声回復法に関する研究
Author(s)	戸井，真智
Citation	
Issue Date	2005-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1918
Rights	
Description	Supervisor:赤木　正人，情報科学研究科，修士

A Speech Dereverberation Method for Improving Speech Intelligibility

Masato Toi (310072)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 14, 2005

Keywords: speech dereverberation, MTF, power envelope restoration, carrier regeneration, intelligibility.

Speech dereverberation is an important issue concerning not only various kinds of speech signal processing such as speech-emphasis for transmission systems and hearing aid systems, but also preprocessing for speech recognition systems. There are several well known inverse filtering methods which can be used to dereverberate the original signal from a reverberant signal in room acoustics. However, for all of these methods the impulse response of the room acoustics must be precisely measured to determine the inverse filtering before the dereverberation. Moreover, the impulse response temporally varies with various environmental factors (temperature, etc.), so the room acoustics have to be precisely measured each time these methods are used. This is a significant drawback with regard to the use of these methods for various speech applications.

On the other hand, Sakata *et al.* proposed a dereverberation speech method based on the modulation transfer function (MTF) concept for without measuring the room impulse response. In this model, the envelopes and carriers are decomposed from speech signal using a constant-bandwidths filterbank and then they are carried out a dereverberation processing in each channel, respectively. However, this model has to be considered the following points: (1) how to design an adequate filterbank to make the best improvement of the restoration accuracy in the power envelope restoration model; and (2) how to develop the carrier reconstruction method to improve the sound quality of the reconstructed sound.

The goal of our work is to construct a blind speech dereverberation method which can restore a speech signal from reverberant speech without using useful prior information such as the impulse response of the room acoustics, and which enables less loss in speech intelligibility caused by reverberation. In this study, a speech dereverberation model is constructed based on Sakata's method, because this method can restore the original speech from a reverberant speech without measuring the room impulse response. Moreover, this method has possibility of improving the speech intelligibility, because this method restores not only speech power envelope but also speech carriers from reverberant speech.

In this model, the reverberant speech signal is decomposed into the envelopes and carriers in N-channels using a filterbank. First, in the envelope dereverberation process, the power envelope inverse filtering method is used to recover the power envelope of original speech from the reverberant envelope. In the previous model, the constant-band N-channel filterbank (N =100) was used and the power envelope inverse filtering was done for a whole duration of speech in each channel. So, to improve the power envelope restoration accuracy, we extend our model to an adaptive time-frequency division processing using a reconstructed filterbank depending on each speech signal. With regard to time-division, a threshold method is used to determine time-segments from a whole duration in each channel. Then, power envelope inverse filtering process is carried out in each time-segment. With regard to frequency-division, the co-modulation characteristics of the target speech are used to determine the adequate bandwidth in the filterbank. We examined the correlation between the power envelopes on channels in a constant narrow-band (40 Hz) filterbank to verify the co-modulation characteristics.

Second, this model employed the PIFM in carrier regeneration process. In this process, a source generator is used to reconstruct the source information of the dereverberated speech from voiced/unvoiced information based on the estimated fundamental frequency (F0), and then the source information is decomposed into the carriers in channels. In the previous model, however, there is no treating with phase information so that sound quality of the reconstructed speech seems to be heard a "buzzy" sound. Thus, in order to remove "buzzy" sounds, the controlling of the group delay, employed by the STRAIGHT, is also adopted to this process. In this study model, the improved method manipulates the group delay of the source signal at more than 3 kHz via random phases.

We evaluated these models by simulation and listening tests. In the simulation, we evaluated with regard to the power envelope restoration in each channel using three measures: correlation, SNR and log-spectrum distortion (LSD). As a result, we found that the amount of the improvement of all evaluations of proposed method is larger than that of previous method. Moreover, in the listening test, we used the Scheffe's method of paired comparison to measure the sound quality and the reverberantness of dereverberated speech perceived by the subjects. We could see that the sound quality of the improved model is better than that of the previous model. However, subjects experienced more reverberantness with the improved model than with the previous model. By adjusting the group delay, sound quality and reverberantness come to a compromise and the dereverberated speech sounds more natural. Finally, we measured the speech intelligibility of the dereverberated speech. We could see that the word intelligibility with the proposed model is higher than that with the previous model and the reverberant version. As a result, it is found that the improvement of the speech quality greatly contributes to the improvement of speech intelligibility. These results showed that the proposed model can accurately restore reverberant speech in both subjective and objective evaluations.