JAIST Repository
https://dspace.jaist.ac.jp/

Title	瞬時振幅を利用したブラインド残響音声回復の可能性の検討
Author(s)	柴野，洋平
Citation	
Issue Date	2007-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/3624
Rights	
Description	Supervisor:鵜木　祐史，情報科学研究科，修士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

An examination of possibility of speech dereverberation using instantaneous amplitude

Yohei Shibano (510045)

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

February 8, 2007

Keywords: Speech dereverberation, Carrier, Instantaneous amplitude, Instantaneous phase.

Speech intelligibility lowers in reverberant environments because the reverberation smears significant features of speech. Speech dereverberation is, therefore, needed in various speech signal processing such as hearing aid systems and preprocessors for ASR systems. The ultimate goal of our work is to construct a blind speech dereverberation method which can restore a speech signal from reverberant speech without measurement of the room acoustics, and which reduces the loss of speech intelligibility caused by reverberation. A speech dereverberation method based on the modulation transfer function (MTF) was proposed by Unoki *et al.* This method consists of power envelope restoration and carrier regeneration processes. In this method, the power envelopes and carriers are decomposed from the reverberant speech using an N-channel narrow-band filterbank and they are redecomposed using adaptive time-frequency division. In power envelope restoration process, the power envelopes are individually restored using MTF-based power envelope inverse filtering. In carrier regeneration process, carrier signals were separately regenerated using the pulse source interpolated frequency modulation (PIFM) method. All regenerated carriers are added together and are passed through all-pass filtering to manipulate phase information by the group delay control. The current dereverberation method contributed to the recovery of reverberant speech with regard to the signal restoration and speech intelligibility. However, since we incorporated F0 estimation method into this dereverberation method to regenerate carrier signals, the regenerated carriers may be affected by accuracy of the estimated F0 from a reverberant speech. In this paper, we try to improve the carrier regeneration process in the MTF-based dereverberation method without using the estimated F0 from the reverberant speech. In the current method, the instantaneous

Copyright © 2007 by Yohei Shibano

amplitudes and phases of channel signals in the filterbank can be directly derived using the Hilbert transform. Moreover, the power envelope of the room acoustics can be also estimated as the sense of restoration of the power envelopes of the channel signals. Therefore, the carrier can be artificially reconstructed using the Hilbert transform relations between the instantaneous amplitudes and phases of them. The aim of this paper is to show the possibility that the improved method can regenerate carriers and then can blindly restore the reverberant speech without using the estimated F0.

We resynthesize a signal from information of the envelope which recovered enough without using the estimated F0. In improved method, reverberation speech is divided, and instantaneous amplitude is restored by power envelope inverse filtering. We estimate instantaneous phase from the restored amplitude, and restore a speech.

Channel signal $x_n(t)$ is defined as AM-PM signal:

$$x_n(t) = s_{x,n}(t)\cos(\omega_{c,n}t + p_{x,n}(t)), \tag{1}$$

where $s_{x,n}(t)$ is the instantaneous amplitude, $p_{x,n}(t)$ is the instantaneous phase, $\omega_{c,n}$ is center frequency of channel, and n is channel number. Here, the instantaneous amplitude $s_n(t)$ and the instantaneous phase $p_n(t)$ are represented as

$$s_{x,n}(t) = \sqrt{x_n(t)^2 + \{\mathrm{Hilbert}(x_n(t))\}^2}, \tag{2}$$

$$\omega_{c,n}t + p_n(t) = \arctan(\mathrm{Hilbert}[x_n(t)]/x_n(t)). \tag{3}$$

The Hilbert transform of $x_n(t)$ is

$$\mathrm{Hilbert}(x_n(t)) = \mathrm{F}^{-1}[-j\mathrm{sgn}(\omega)X_n(\omega)], \tag{4}$$

where $X_n(\omega)$ is the Fourier transform of $x_n(t)$. The instantaneous amplitude is estimated by low-pass filtering the restored envelope. The small high frequency ingredients are included in the envelope, because LPF is not an ideal filter. Since the envelope is estimated by low-pass filtering the instantaneous amplitude, we may estimate the instantaneous amplitude by low-pass inverse filtering the envelope. Therefore, We can estimate the instantaneous amplitude of a restored signal $\hat{s}_{x,n}(t)$. This paper assumes that the instantaneous amplitude of original signal was estimated from an observed signal. If the instantaneous phase is known, we can restore the signal. In this paper, we estimate instantaneous phase of original signal from instantaneous amplitude.

In the frequency domain, we can estimate the phase from the amplitude, when a signal is minimum phase. The signal that real part is positive is minimum phase signal. The minimum phase signal is causality signal. Fourier transform of minimum phase signal is

$$\begin{aligned} X(\omega) &= s(\omega)\cos(p(\omega)) + js(\omega)\sin(p(\omega)) \\ &= s(\omega)\exp(jp(\omega)), \end{aligned} \tag{5}$$

where $s(\omega)$ is amplitude, and $p(\omega)$ is phase spectra. There is Hilbert transform relations between the real part and the imaginary part. In the logarithm representation, the amplitude and phase can be divided into the real part and the imaginary part.

$$\log[X(\omega)] = \log[s(\omega)] + jp(\omega) \tag{6}$$

In the minimum phase signal, the inverse Fourier transform of $\log[X(\omega)]$ is causal signal. In the case, Hilbert transform relations between the logarithm amplitude and phase exist. Therefore, we can estimate the phase from the amplitude.

In the time domain, we estimate phase from the amplitude, from the analogy between the time domain and frequency domain. We ignore center frequency of filterbank $\omega_{c,n}$. The analytic signal $x_{a,n}(t)$ is defined by channel signal that divided speech $x_n(t)$ and the Hilbert transform.

$$\begin{aligned} x_{a,n}(t) &= x_n(t) + j\mathrm{Hilbert}[x_n(t)] \\ &= s_{x,n}(t)\cos(\omega_{c,n}t + p_{x,n}(t)) + js_{x,n}(t)\sin(\omega_{c,n}tp_{x,n}(t)) \\ &= s_{x,n}(t)\exp(jp_{x,n}(t))\exp(j\omega_{c,n}t) \end{aligned} \tag{7}$$

There is the Hilbert transform relations between the real part and the imaginary part. The logarithm of $x_{a,n}(t)$ is represented

$$\log[x_{a,n}(t)] = \log[s_{x,n}(t)] + j(p_{x,n}(t)). \tag{8}$$

This argument in time domain is the same as a frequency domain. The concept of this paper is, if there is Hilbert transform relations between the logarithm instantaneous amplitude and instantaneous phase, we can restore a signal form instantaneous amplitude. In this paper, we estimate instantaneous phase using

$$\hat{p}_{x,n}(t) = \mathrm{Hilbert}(\log|s_{x,n}(t)|). \tag{9}$$

We carried out 100 simulations to evaluate the improved carrier regeneration process with regard to time and frequency division. The speech signals were 10 Japanese sentences uttered by 10 speakers (5 males and 5 females) in the ATR-database datasets. Sampling frequency is 20 kHz. Constant-bandwidths filterbank was used. In these simulations, SNR and LSD were used as evaluation measure are used to show the improvement in restoration with this improved model. The result showed that Mean of SNR of restored speech is high to be about 24.0 dB. Mean of LSD of restored speech is low to be about 0.1 dB. Therefore, we restored the speech from the instantaneous amplitude.

In this paper, We improved carrier regeneration process in the MTF-based speech dereverberation method for restore a speech signal from reverberant speech without measurement of the room acoustics. Depending on the estimation of instantaneous amplitude of original signal by the previous work, We restore the speech from the obtained instantaneous amplitude. We have carried out 100 simulations (10 Japanese sentences) of 10 speakers (5 males and 5 females in the ATR-database datasets) of speech signals to evaluate the improved. The result showed that mean of SNR is high to be about 24.0 dB, and mean of LSD is low to be about 0.1 dB. When instantaneous amplitude, We can restore reverberation speech. The instantaneous amplitude may be restored by operating power envelope inverse filtering only in low level of spectrums, and narrowing a bandwidth when We divide reverberation speech. If instantaneous amplitude was restored, We can restore reverberation speech. Therefore, reverberation speech may be restored without using F0. We showed the possibility that the improved method can regenerate carriers and then can blindly restore the reverberant speech without using the estimated F0. In our future work, we will evaluate this improved method by using reverberant speech.