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

Title	複素ケプストラム分析を用いた残響音声の基本周波数推定に関する研究
Author(s)	細呂木谷，敏弘
Citation	
Issue Date	2007-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/3603
Rights	
Description	Supervisor:鵜木　祐史，情報科学研究科，修士

JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

A study on a fundamental frequency estimation for reverberant speech based on the complex cepstrum analysis

Toshihiro Hosorogiya (510091)

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

February 8, 2007

Keywords: Fundamental frequency, Reverberation, Complex cepstrum analysis, Source-filter model.

An extraction of the fundamental frequency (F0) of a target speech is a very important issue for not only speech analysis/synthesis but also various speech signal processing such as speech separation and speech dereverberation. Many studies on estimating the F0 have been done in about 50 years and then many methods have been proposed. Currently, a few methods for precisely estimating F0 in clean environment are established. However, these methods are affected by noise and/or reverberation when these are applied to various speech signal processing in real environments. Therefore, a robust F0-estimation method in real environments is required. Recently, some robust methods for estimating the F0 from a noisy speech have been proposed. These methods can precisely estimate the F0 in noisy environments. However, there are no robust studies for reverberant environments and there are no proposed methods for estimating the F0 from a reverberant speech. Therefore, this paper proposes a robust estimation method from a reverberant speech. If we can correctly estimate F0 from a reverberant speech, we can use this F0 for dereverberation, source separation, and other speech signal processings. For example, it is known that most speech recognition systems are extremely affected by reverberation, a cor-

Copyright © 2007 by Toshihiro Hosorogiya

rect F0 can contribute to improving correct recognition rate by using F0 for subdividing words or phrases.

In this paper, traditional F0 estimation methods are evaluated with regard to robustness in reverberant environments. To investigate robustness of traditional methods, ten traditional methods were used. These methods were methods using periodicity in time domain (Auto-correlation method, AMDF method, LPC residual method, and ACMWL method), and methods using harmonicity in frequency domain (STFT-comb filtering method, SHS method, Liftering method, cepstrum method, TEMPO, and PHIA method).

The speech database used in the evaluation is the database which has speech and EGG (electro glottal graph) data. The data used in this evaluation consists of four Japanese utterances spoken by each twelve male and female speakers.

The addition of the reverberation is obtained by convoluting speech with a reverberant impulse response which has non-minimum phase and exponential attenuate characteristics. Evaluation measure used in this evaluation are a correct rate and a SNR. The correct rate indicates a robustness to reverberation. The SNR indicates an accuracy of the estimated F0.

As the results, correct rates of all methods are drastically reduced as the reverberation time increases. Especially, when reverberation time is 2 s, correct rates of all methods are less than 50%. Hence, it is found that these methods cannot work well in reverberant environments. The accuracy of the cepstrum method is superior to the those of the other methods excluding a clean case.

The cepstrum analysis is a homomorphic analysis to convolution. This analysis can treat a direct sound and a reflect sound to be the same signal. Therefore, it is supposed that the cepstrum method is not affected much by reverberation.

Considering this evaluation, this paper proposed an F0 estimation method based on the complex cepstrum analysis. This method is composed of two blocks. The first block extracts the speech source information from the reverberation speech by using complex cepstrum anslysis. When the source filter model is employed, a source information related to vocal fold and a filter information related to vocal tract are separated in each high quefrency and low quefrency part. Therefore, if only the component on low quefrency part is removed by lifter, only speech source information necessary to estimating F0 can be extracted. This operation eliminates the component of the reverberation on lower quefrency parts, and to removing a dominant component of the reverberation simultaneously.

The second block estimates the F0 from remaining speech source information using periodicity or harmonicity. The various methods can be applied as the processing of the second part either treat periodicity or harmonicity. In this paper, auto-correlation method in the time-domain is used as a method of treating periodicity, and comb-filtering method of the amplitude spectrum is used as a method of treating periodicity.

To evaluate effectiveness of the proposed method, a correct rate and SNR of the estimated F0 using the proposed method is compared with these of ten traditional methods. Here, the same dataset and evaluation measure are used. The result shows that the correct rate of the estimated F0 by the proposed method using comb-filtering is not reduced much than the other methods as the reverberation time increases . Therefore, the proposed method is more robust than the other methods for reverberant environments, and a speech source information extracted by the complex cepstrum analysis is found as an effective feature for a estimation F0 from a reverberant speech.

As the result of the evaluation, it is supposed that the effect of the reverberation in the analysis frame can be mostly removed by the proposed method. However, in this method, we do not consider the effect that previous analysis frame information gives present analysis frame because the length of reverberant impulse response is longer than that of analysis window. Therefore, this report considered the usage of processing based on more long window length than the reverberant impulse response.

First of all, the main feature of reverberation which affected a speech was investigated. The result shows that a non-minimum phase component of reverberation affects an estimation F0 more than a minimum phase component. Therefore, this report considered to remove this non-minimum component by using a CMN(Cepstrum Mean Subtraction). However, this method did not affect, because it is difficult to estimate the CMN of a non-minimum phase component.