Title	雑音環境における基本周波数推定とこれを用いた雑音抑圧に関する研究
Author(s)	石本，祐一
Citation	
Issue Date	2000-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1362
Rights	
Description	Supervisor:赤木 正人，情報科学研究科，修士

A study on a fundamental frequency estimation method and noise reduction in noisy environment

Yuichi Ishimoto

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

February 15, 2000

Keywords: fundamental frequency estimation, noisy environment, comb filter, noise reduction.

1 Introduction

Some speech recognition systems are of practical use. However, their performances reduce in noisy environments. Especially hearing aid users feel noisiness. Addtionally, background noises interfere understanding conversations. Extraction of speech waveforms and features in noisy environments are needed for use of speech signal processing as speech analysis/synthesis and speech segregation. Kawahara et al. proposed a fundamental frequency (F0) estimation method (TEMPO2) using fixed points of frequency to instantaneous frequency map[1]. Although TEMPO2 can estimate F0s for clean speech accurately, it cannot work in noisy environments. On the other hand, although another method using comb filters proposed by Unoki et al[2] can estimate F0s in noisy environment, the estimated F0s are not accurate enough. This paper proposes a robust and accurate F0 estimation and noise reduction methods, which consists of a comb filter with controllable pass bands, TEMPO2, and a F0 estimation method using comb filters.

2 Outline of the proposed method

An overview of the proposed method is shown in Figure 1.

A mixed signal $x(t)$ with a harmonic target signal $s(t)$ and noise $n(t)$ is observed by a single microphone. In the F0 estimation section,

1. F0 ($\bar{F}0$) of the target signal are estimated roughly by the method using comb filters [2] from the observed mixed signal $x(t)$.

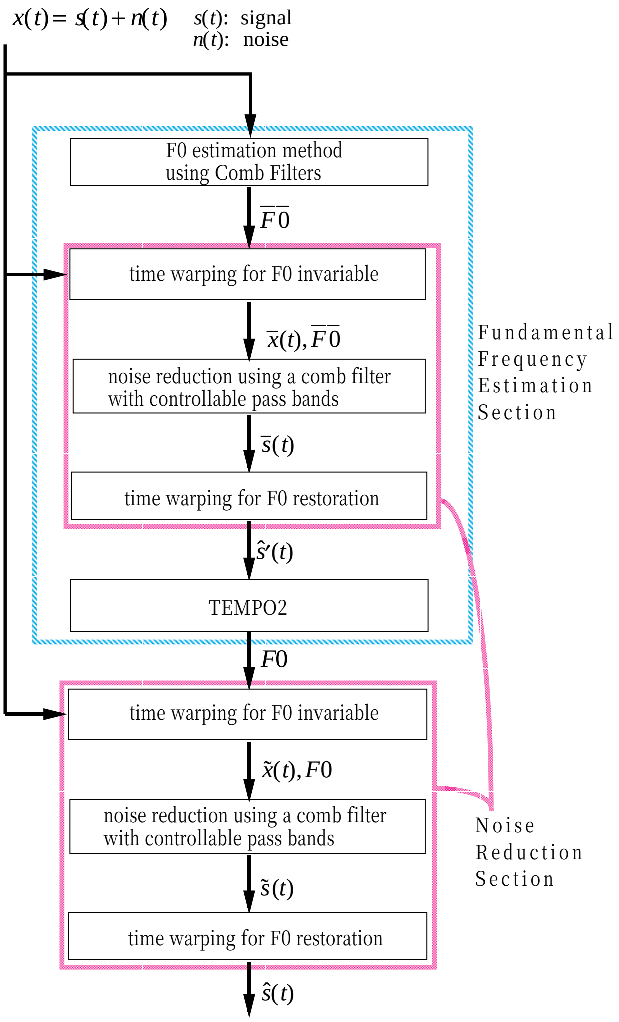

Figure 1: Outline of the proposed method

2. Noise reduction is achieved using a comb filter with controllable pass bands, whose center frequencies are multiples of the estimated F0 roughly and whose band-widths are controlled not to reduce harmonic components of speech.

   When reducing noise, time warping of the mixed signal for fixing F0 is performed to decrease errors for noise reduction.

3. TEMPO2 [1] is applied to the noise-reduced signal $\hat{s}'(t)$ for accurate F0 estimation.

Furthermore, in the noise reduction section,

4. Noise reduction is achieved using the comb filter whose center frequencies are the estimated F0s accurately.

Then, the target signal $\hat{s}(t)$ is obtained.

3 Noise reduction using the comb filter with controllable pass bands

A mixed signal $x(t)$ is defined as follows,

$$\begin{aligned} x(t) &= s(t) + n(t) & (1) \\ &= \sum_m a_m e^{j(m\omega_0(t)t+\theta_m)} + \sum_k b_k e^{j(\omega_k t+\theta_k)} & (2) \\ & \quad (\omega_0(t) = 2\pi/T(t)), \end{aligned}$$

where $s(t)$ is a harmonic signal with a fundamental period $T(t)$ and $n(t)$ is a noise signal. Assuming $T(= 2\pi/\omega_0)$ is an invariable value of $T(t)$, a signal $g(t)$, which is the subtracted signal shifted $x(t)$ to $\pm T$ in time from $x(t)$, is represented as follows.

$$\begin{aligned} g(t) &= \frac{2x(t) - x(t-T) - x(t+T)}{4} & (3) \\ &= \sum_k b_k e^{j(\omega_k t+\theta_k)} \sin^2 \frac{\omega_k}{\omega_0}\pi. & (4) \end{aligned}$$

Fourier transform $G(\omega_k)$ of $g(t)$ is

$$G(\omega_k) = N(\omega_k) \sin^2 \frac{\omega_k}{\omega_0}\pi, \tag{5}$$

where $N(\omega_k)$ is Fourier transform of $n(t)$. Then, the noise spectrum $N(\omega_k)$ is represented as

$$N(\omega_k) = G(\omega_k)/\sin^2 \frac{\omega_k}{\omega_0}\pi. \tag{6}$$

The target signal $s(t)$ can be estimated by subtracting the noise signal $n(t)$, which is inverse Fourier transform of $N(\omega_k)$, from $x(t)$. However, if ω_k/ω_0 is an integer, the noise spectrum $N(\omega_k)$ in equation (6) is infinity. For practical applications, a parameter ϵ is established.

$$\tilde{N}(\omega_k) = \begin{cases} G(\omega_k)/\sin^2 \frac{\omega_k}{\omega_0}\pi & \left|\sin \frac{\omega_k}{\omega_0}\right| \geq \epsilon \\ G(\omega_k) & \left|\sin \frac{\omega_k}{\omega_0}\right| < \epsilon. \end{cases} \tag{7}$$

This means a comb filter that can control pass bands by a value of ϵ.

Temporal fluctuations of fundamental periods in real speech cause errors of the estimated noise $\tilde{n}(t)$, because fundamental period is assumed an invariable value in equation (3). Thus, time warping of the mixed signal for fixing the fundamental periods is performed.

4 Fundamental frequency estimation in noisy environment

For applications of noise reduction using the comb filter with controllable pass bands, F0 of the target signal must be estimated in noisy environments. First, performances of

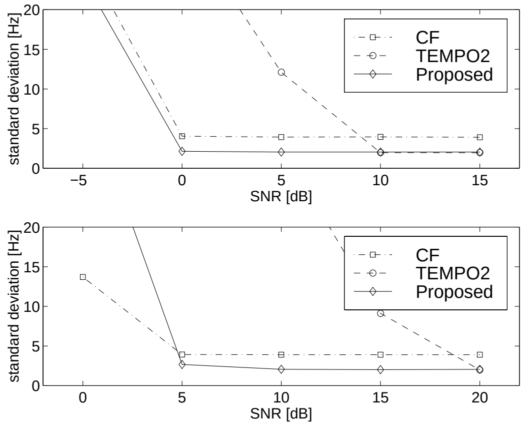

Figure 2: relationship between estimation accuracies and SNRs : white noise (upper), pink noise (lower)

well-known F0 estimation methods for noisy signals are investigated. When noises were white and pink noises, the results showed that,

- If the signal-to-noise ratio (SNR) of the noisy signal is high, TEMPO2 is the most accurate. However if the SNR is low, its performance reduces extremely.
- If the SNR of the noisy signal is low, the method using comb filters is the most accurate. However if the SNR is high, its performance does not reach that of TEMPO2.

A robust and accurate F0 estimation method is constructed using noise reduction by the comb filter with controllable pass bands, described above. Algorithm of the proposed method consists of the comb filter with controllable pass bands and two types of F0 estimation methods, which are TEMPO2 and the method using comb filters, as described at the F0 estimation section in Figure 1.

Figure 2 shows relationship between estimation accuracy (standard deviations of estimation errors) and SNRs. Figure 2 indicates that the proposed method is a robust and accurate method, which can estimate accurate F0s in noisy environments larger than 5dB SNR.

5 Simulations of noise reduction

Extraction of the target signals described in Table 1 is performed using the proposed method.

Noises that SNRs are ranged from 0 to 20 dB in 5 dB steps added the target signals. SNR, spectral distortion (SD) and auditory-oriented spectral distortion (ASD)[3] are used as evaluation measures. Simulated results show that the proposed F0 estimation and noise reduction methods in noisy environments are effective when the noises are white and vowels. The method can increase SNR by 5-7dB. However, when the noise is pink,

Table 1: experimental data

Sim. No.	target signal	noise
1	/a/ /i/ /u/ /e/ /o/ /aoi/ (Speeker mht,mau,fsu,fkn : ATR speech database)	white noise
2	/a/ /i/ /u/ /e/ /o/ /aoi/ (mht,mau,fsu,fkn)	pink noise (60〜6000kHz)
3	/a/ (fsu)	/a/ (mht)

the method must be improved to reduce noises which remain harmonic bands of target signals, and not to eliminate their comonents except harmonic bands.

6 Conclusion

This paper proposed robust and accurate F0 estimation and noise reduction methods that consists of a noise reduction method using the comb filter with controllable pass bands, TEMPO2 which is the accurate F0 estimation method, and the robust F0 estimation method using comb filters. The proposed method can estimate F0s as accurate as clean speech in noisy environments larger than 5dB SNR. Additionally,this paper proposed a noise reduction method using comb filter whose center frequencies are the estimated F0s accurately. The method can increase SNR by 5-7dB.

References

[1] Hideki Kawahara, Parham Zolfaghari, Alain de Cheveigné, Roy D. Patterson, Source Information Extraction using Fixed Points of a Freqency to Instantaneous Frequency Map, IEICE,Tech.Rep., SP99-40, July 1999.

[2] Masashi Unoki, Masato Akagi, A Method of Extracting the Harmonic Tone from Noisy Signal Based on Auditory Scene Analysis, Translated from IEICE, vol.J82-A, No.10, pp.1497-1507, 1999.

[3] Mitsunori Mizumachi, Masato Akagi, An Objective Distortion Estimator for Hearing Aids and its Application to Noise Reduction, EUROSPEECH'99, pp.2619-2622, 1999.