

Pitch Detection of Duet Song using Double Comb Filters

Yoshiaki Tadokoro* and Michiru Yamaguchi*

Abstract-This paper proposes a new method for the pitch detection of a duet song for transcription. The previous pitch detection methods that are considered most for musical instrument sounds but few for sung songs are based on the extraction of the pitch frequencies. But the extraction of the pitches is difficult since the musical sound has many harmonic components. The principle of our method is the elimination of the pitch and its harmonic frequencies. This is performed simply by a comb filter. To adapt the fluctuation of song pitches, we use a double comb filter(DCF). Using two cascade connected twelve DCFs that are connected in parallel and are corresponding to each tone in one octave, we can detect the pitches of a duet song.

1 Introduction

Musical transcription is necessary in musicology field and also one of the interesting targets as application of machine perception and artificial intelligence [1]-[3]. The musical sound is composed of a fundamental (pitch) frequency corresponding to each tone and its harmonic frequencies. The pitch detection plays the most important role for transcription. The previous pitch detection methods that are considered most for musical instrument sounds but few for sung songs are based on the extraction of the pitch frequencies[4][5]. But the extraction of the pitches is generally difficult, especially for polyphony, since the musical sound has many harmonic components, and so complicated signal processing is necessary. To settle this problem, we proposed a new pitch detection method that is based on the elimination of the pitch and its harmonic frequencies [6]-[8]. This is performed simply by a comb filter ($H(z) = 1 - z^n$). By using the cascade connected twelve comb filters corresponding to each tone in one octave and detecting zero-output of these comb filters, we can detect the pitches for polyphonic instrument sounds. But we could not apply this method directly for sung songs, especially for polyphonic songs, because it has the fluctuation of pitch frequencies and strong formant components comparing with pitch ones.

In this paper, to overcome these difficulties, we use a double comb filter (DCF) that has the some extended notch frequency width and connect twelve

DCFs in parallel. We show that we can detect the pitches of duet songs by connecting two parallel 12 DCFs in cascade.

2 Characteristics of Sung Song

Figure 1 shows the waveforms and the DFT results of a tone *A* of octave 3 (*A3*) for (a) piano sound, and (b) male's "Ka" sound in Japanese, respectively. They are composed of a fundamental (pitch) frequency and many harmonic frequencies. Comparing with the instrument sound, the song sound has the fluctuation of the pitch frequency, strong components at formant frequencies and consonant components. These characteristics may make the pitch detection of a song difficult.

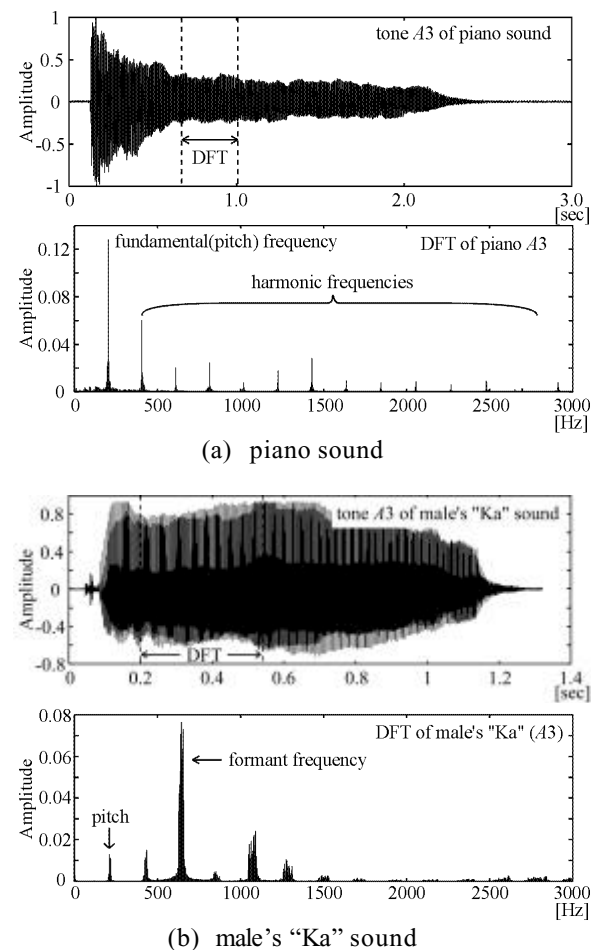


Fig.1 Waveforms and DFT results of tone *A* of octave 3 (*A3*) for (a) piano sound and (b) male's "Ka" sound.

*Toyohashi University of Technology
Hibarigaoka 1-1, Tempaku, Toyohashi, 441-8580 JAPAN
Phone:(81)532-44-6754 Fax:(81)532-44-6757
E-mail:tadokoro@signal.tutics.tut.ac.jp

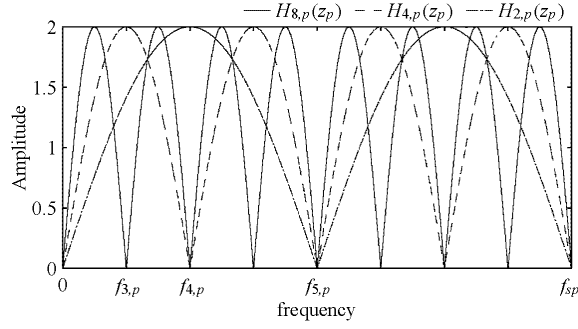


Fig.2 Amplitude frequency characteristics of comb filters $H_{q,p}(z_p)$ ($q=2,4,8$).

3 Principle of Pitch Detection using Comb Filters

3.1 Comb Filter

The transfer function of a comb filter can be written by

$$H_{q,p}(z_p) = 1 - z_p^{-q}, \quad (1)$$

$$p = 1, 2, \dots, 12, \quad q = 2, 4, 8$$

where p shows one of 12 tones in one octave and q is the order of the comb filter.

The amplitude frequency characteristics of $H_{q,p}(z_p)$ are illustrated in Fig.2. The sampling frequency f_{sp} for the tone name p of the octave 3 is determined by $8 \times f_{3,p}$, where $f_{3,p}$ shows the fundamental frequency of the tone p in the octave 3.

Actually, we realize the comb filter by the over-sampling method for the cascade connection of some comb filters, since we cannot connect several comb filters having different sampling frequencies directly. That is, we use a higher sampling frequency f_s instead of f_{sp} when we sample the musical sound to digitize. We use the unit delay z^{-1} ($1/f_s$) instead of the unit delay z_p^{-1} ($1/f_{sp}$) and we approximate Eq.(1) by

$$H_{8,p}(z) = 1 - z^{-n_p}, \quad (2)$$

where

$$n_p = 8 \cdot \lceil f_s / f_{sp} \rceil \cdot [\text{an integer by rounding}] \quad (3)$$

Practically, we use $f_s = 44.1 \text{ kHz}$ and so the maximum error of the approximate f_{sp} is about 0.18 %. By this over-sampling method, each comb filter has the same sampling frequency f_s , and so we can connect several comb filters in cascade directly.

3.2 Pitch Detection from Zero-output of Comb Filters

From the frequency characteristics of Fig.2, we can realize that $H_{8,p}(z_p)$ can eliminate the pitch frequency of the tone p of the octave 3 and its harmonic frequencies. Therefore, if the output of $H_{8,p}(z_p)$ is zero when the musical sound is entered

into $H_{8,p}(z_p)$, then we can know that the pitch of the sound is the tone p . Furthermore, we can know that $H_{4,p}(z_p)$ can eliminate the fundamental frequency $f_{4,p}$ and its harmonic frequencies over the octave 4, but does not eliminate $f_{3,p}$ and its odd number harmonic frequencies. Similarly, $H_{2,p}(z_p)$ can eliminate $f_{5,p}$, but $f_{3,p}$ and $f_{4,p}$. From this feature, we can detect the octave number of the musical sound.

3.3 Effect of Double Comb Filters

In our method, the pitch detection can be made by the detecting the zero-output of the comb filters. But, owing to the fluctuation of the pitch frequency and components other than its harmonic components, the output of the comb filter cannot show the complete zero. We determine the pitch from the comb filter showing the most smallest output by comparing outputs of comb filters. Therefore it is desired that the output of the comb filter that is for elimination of the pitch and its harmonic frequencies becomes almost zero.

For adapting the fluctuation of the pitch frequency of the sung song, we consider a double comb filter $H_{q,p}^2(z)$, connected two comb filters $H_{q,p}^+(z)$ and $H_{q,p}^-(z)$ in cascade, representing in the following form :

$$H_{q,p}^+(z) = 1 - z^{-np+\Delta np} \quad (4)$$

$$H_{q,p}^-(z) = 1 - z^{-np-\Delta np} \quad (5)$$

$$H_{q,p}^2(z) = H_{q,p}^+(z) \cdot H_{q,p}^-(z) \quad (6)$$

This double comb filter can have some extended width of the notch frequency as shown in Fig.3. Figure 4 shows the effect of the double comb filter that has the smaller zero-output than in a single comb filter.

4 Pitch Detection Method of Duet Song using Double Comb Filters

To detect the pitches of the musical instrument sound, we used the twelve cascade connected comb filters shown in Fig.5, and determined the first pitch of a polyphony from the comb filter showing almost zero-output. Then the comb filter showing the zero-output is moved to the first stage before the comb filter $H_{8,1}(z)$, and again we find the comb filter showing zero-output to detect the second pitch. By this method, we can detect the pitches of any polyphony sound if each tone has different pitch. But for the sung song, we could not detect their pitches correctly using this method, because of the feature of the sung song described in 2.

For the solo and duet song of octaves 3 and 4, we consider the implementation showed in Fig.6 to detect these pitches of the songs. In this configuration, the outputs are obtained through only two double comb filters and the total number of double comb filters is 78.

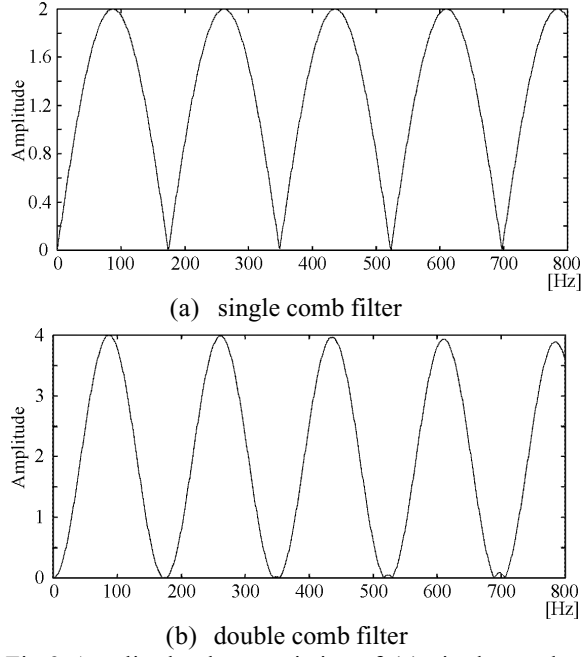


Fig.3 Amplitude characteristics of (a) single comb filter and (b) double comb filter for sound $F3$ ($n_p=253, \Delta n_p=3$).

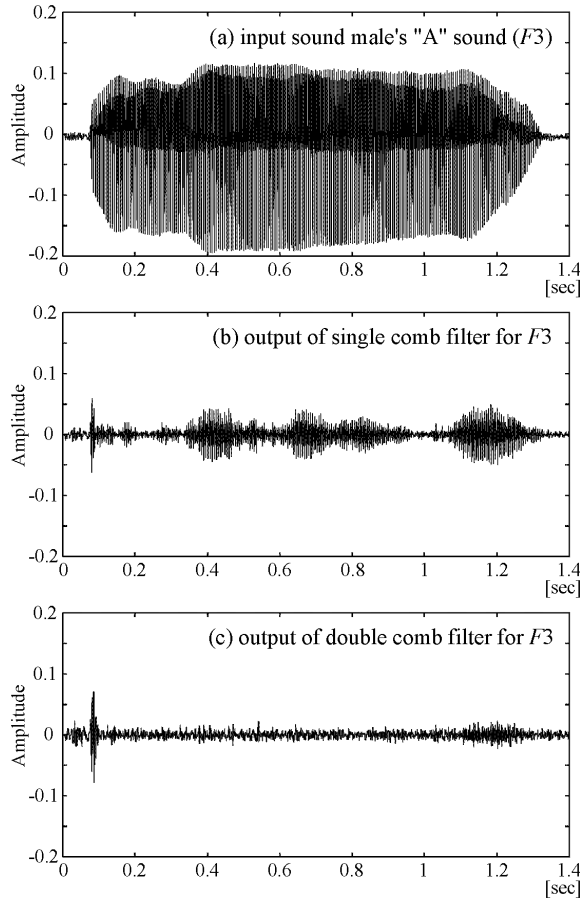


Fig.4 Effect of double comb filter. (a) input sound male's "A" sound ($F3$), (b) output of single comb filter and (c) output of double comb filter ($n_p=253, \Delta n_p=3$).

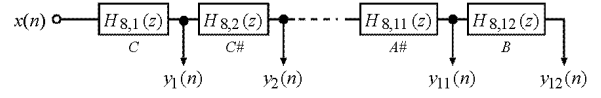


Fig.5 Cascade implementation using single comb filters for the pitch detection of musical instrument sounds.

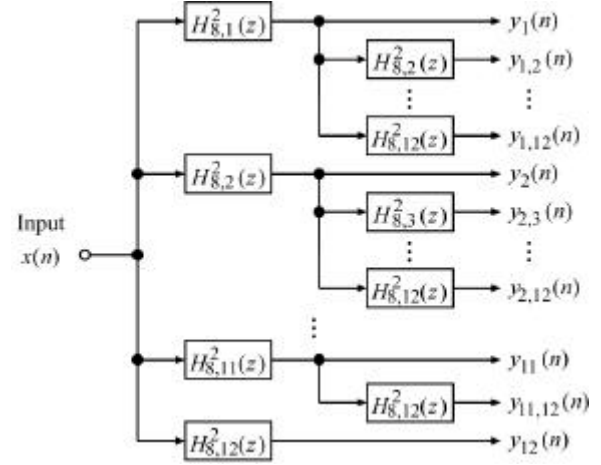


Fig.6 Parallel implementation using double comb filters for the pitch detection of solo and duet songs.

5 Simulation Results

Using the pitch detection system of the duet song shown in Fig.6 ($\Delta n_p=3, p=1, \dots, 12$), we processed the duet song of the score written in Fig.7(a). The input waveform of Fig.7(b) is sampled by the sampling frequency $f_s=44.1\text{kHz}$ and filtered by a lowpass filter that is a six-order butterworth-type filter with the cutoff frequency $f_c=500\text{ Hz}$. We carried out the process of the pitch detection every 2200 sample data (about 50 ms). Figure 8 shows the results of the pitch detection. It is realized that we could obtain almost correct results. The difference between the output values of the comb filters corresponding to the song pitches and other output ones was almost three or four times.

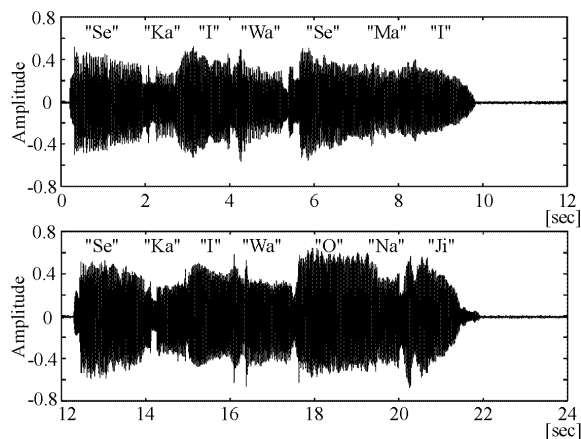
6 Conclusions

We proposed a new method for the pitch detection of the duet song. The method is based on the double comb filters and realized by two cascade connected 12 double comb filters in parallel connection. The number of double comb filters is 78. From the simulation results, it was clear that we could obtain the correct pitch detection by the proposed method. But this method uses many double comb filters for triple and more polyphonic songs.

As the next step, for adapting more polyphonic songs, we would like to consider the possibility of the cascade connection method shown in Fig.5 that is used for the musical instruments. The main reason that we cannot use the implementation of Fig.5 for songs is that the frequency characteristic of Fig.5 has a large amplitude in the range from 50 to 120 Hz. We must improve this frequency characteristic using something like a bandpass filter.



(a) the score used in simulation



(b) the waveform of the sound

Fig.7 (a) The score used in simulation and (b) the waveform of the sound.

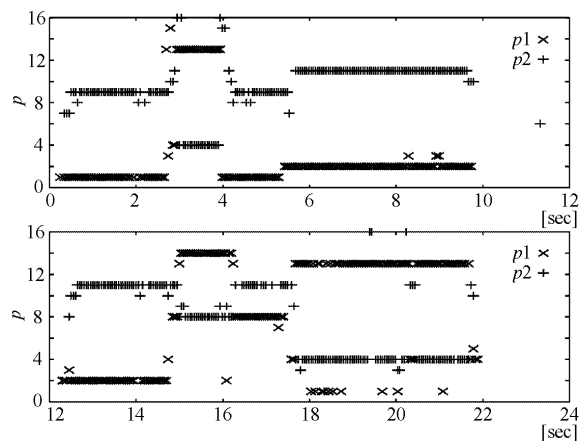


Fig.8 Pitch detection results for the data of Fig.7.

References

- [1] M.Piszczalski and B.A.Galler, "Automatic music transcription," *Computer Musical Journal*, vol.1, pp.24-31, 1977
- [2] C.Roads, "Research in music and artificial intelligence," *ACM computing surveys*, vol.17, pp.163-190, 1985
- [3] W.J.Pielemeier, G.H.Wakefield and M.H.Simon, "Time-frequency analysis of musical signals," *Proc. of the IEEE*, vol.84, no.9, pp.1216-1230, 1996.
- [4] T.Shirado and M.Yanagida, "Extraction of fundamental frequencies from duet sounds," *IEICE Trans.*, vol.E80-A, no.5, pp.912-915, 1997
- [5] H.Katayose and S.Inoguchi, "Intelligent music transcription system," *Trans. of Artificial Intelligence of Japan*, vol.5, pp.59-66, 1990
- [6] T.Miwa and Y.Tadokoro, "Musical pitch estimation and discrimination of musical instruments using comb filters for transcription," *MWSCAS'99*, MP1.4, 1999
- [7] T.Miwa, Y.Tadokoro and T.Saito, "Musical pitch estimation of different sounds using comb filters for transcription," *ICSPAT'99*, p.209(CD-ROM), 1999
- [8] Y.Tadokoro and T.Miwa, "Musical pitch and instrument estimation of polyphony using comb filters for transcription," *CSCC'00, Advances in Physics, Electronic and Signal Processing Application*, pp.315-319, 2000