

Multi pitch estimation by using modified IIR Comb Filters

Mikel Gainza¹, Bob Lawlor², Eugene Coyle³

¹ Digital Media Centre, Dublin Institute of Technology, Dublin, IRELAND

E-mail: *mikel.gainza@dit.ie*

² Department of Electronic Engineering, National University of Ireland, Maynooth, IRELAND

E-mail: *rlawlor@eeng.may.ie*

³ Faculty of Electronic Engineering, Dublin Institute of Technology, Dublin, IRELAND

E-mail: *eugene.coyle@dit.ie*

Abstract - A technique for detecting the pitches of a polyphonic signal is presented. The system utilises modified IIR comb filters, which are generated to ensure that a null (stop band notches) exists at multiples of note frequencies, and that a very flat pass band is present in the remain of the spectrum. Thus, the signal spectrum is not distorted after applying the filters to the audio signal, which is the case when using FIR comb filters. The presented approach improves upon an existing multi pitch detection model based on an FIR comb filter framework.

Keywords – Comb filters, Pitch detection

1. INTRODUCTION

Pitch detection is the core of music transcription systems, and despite the extensive research that has been undertaken in recent years, pitch estimation still remains an open field of research. There are several approaches that deal with the pitch detection problem in monophonic recordings as cepstrum [1, 2], comb filters [3] or time domain autocorrelation methods [4]. Recently, other approaches have also been utilised for polyphonic recordings as auditory models [5], which uses implicit knowledge of the human pitch perception, Independent Component Analysis [6], or comb filters systems [7, 8].

The presented method uses Comb filters to perform the pitch detection process. Comb filters are quite simple to implement and have numerous applications such as pitch detection, music separation [9], speech enhancement [10, 11], audio effects [12] or interference removal in EEG signals [13].

Moorer implemented a pioneering technique that utilises FIR comb filters for the purpose of detecting pitches in monophonic speech recordings [3], and also for detecting guitar chords [14], by observing at the delay which produces a minimum in the output signal. In [7], Miwa also utilise FIR comb filters for music transcription. He built 12 comb filters, one for each note of the octave 3, where the first null (dip) of each FIR comb filter matches exactly the frequency of a different note of the third octave. Then, all the filters are connected in cascade, and the filter which produces a zero output represents the detected note. In [8], to avoid the signal amplitude alteration caused by applying FIR comb filters in cascade, the authors use a bank of parallel FIR comb filters (one per note in the third octave). However, in

the context of rich polyphonic audio, this can lead to spurious pitch detections, since the magnitude of the FIR comb filter frequency response is not flat in between the filter nulls. In order to avoid that amplitude distortion, the proposed multi pitch estimator utilizes a modified comb filter that has a very flat pass band between the nulls. The same filter has been utilised in [13] to remove harmonic interferences.

Section 2 focuses on the comb filter theory. In section 3 the main characteristics of the proposed pitch detection system are described. Some results which validate the approach are shown in section 4 and finally, some conclusions and further work are discussed in section 5.

2. COMB FILTERS THEORY

2.1. FIR comb filters

By using FIR comb filters, the comb spectral shape can be obtained by summing an input signal $x[n]$ with a delayed version of the same signal $x[n]$ [15]. The FIR comb filter transfer function is as follows:

$$H(z) = 1 + g * z^{-D} \quad (1)$$

where g is a factor which scales the gain of the filter between $1+g$ and $1-g$, and D is the delay in samples.

The comb effect results from phase cancellation and summation between the delayed and undelayed signal. This can be appreciated in Fig. 1 where the magnitude responses of a filter with $g = -1$, sampling rate $f_s = 44100$ Hz and a delay $D = 4$ is depicted.

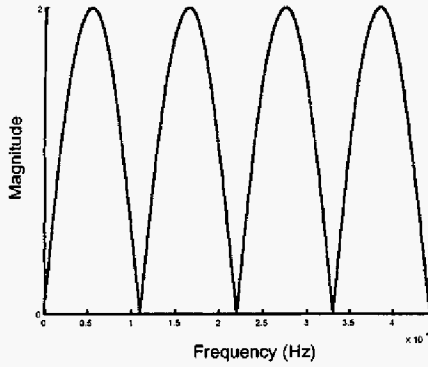


Fig. 1. FIR comb filter magnitude response using $g=-1$, $D=4$ and $f_s=44100$ Hz.

From Fig. 1, it is apparent that at frequencies:

$$n * \frac{f_s}{D} \leq f_s \quad (2)$$

where n is an integer, the delay D causes a 180 degree shift between the original and delayed signal causing cancellation, which produces nulls in the filter magnitude response at frequencies: 11025, 22050, 33075 and 44100 Hz.

2.2. Modified IIR comb filter.

In order to avoid the signal amplitude alteration caused by the FIR comb filter, the authors in [13] introduced a zero for every pole in the frequency response of equation 1. Thus, the frequency response of the modified comb filter becomes:

$$H(z) = \frac{1 - z^{-D}}{1 - \rho^D z^{-D}} \quad (3)$$

where $0 < \rho < 1$ is a stability condition

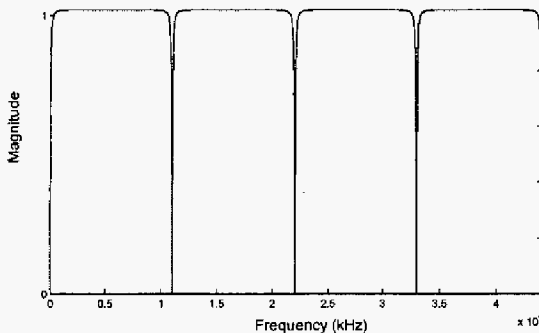


Fig. 2. IIR comb filter magnitude response using $\rho=0.99$, $D=4$ and $f_s=44100$ Hz.

The frequency response of the modified IIR filter is depicted in Fig. 2, where it can be appreciated that the filter pass-band is very flat.

3. PITCH DETECTION SYSTEM

The proposed pitch detection system is based on the bank of parallel FIR comb filters system that was proposed in [8]. The authors generated 12 FIR comb filters, one for each note of the octave, where the first filter null matches accurately the frequency of a different note of an octave. (see Fig. 3). Then, all the filters are applied to the input signal, and the filter which produces an amplitude minimum represents the detected note. Next, the existence of other notes in the audio signal is investigated. This is performed by iteratively connecting the output of the filter that has produced the minimum with the input of the parallel comb filters system, and the same filtering process is repeated again until all the notes have been extracted.

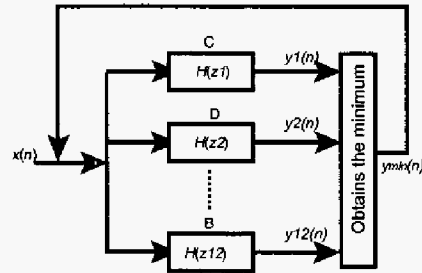


Fig. 3. Parallel comb filters system

To perform the filtering process, different filters were investigated. In [8], the authors used standard FIR comb filters. However, the amplitude filter response of those filters substantially alters the remaining components of the spectrum. Considering Fig. 4, where an audio signal with notes C4 and E4 playing together is depicted, it can be appreciated that the FIR comb filter, which is depicted using a dashed line, filters the harmonics of C4, but also distorts the harmonics of E4.

In order to obtain a flat response between the filter nulls, the filters introduced in section 2.2 [13] can be utilised, which significantly preserve the remaining spectrum after filtering. This is crucial in the context of rich polyphonic transcription. However, musical notes are usually represented in the frequency domain by more than one frequency bin. This is the case when the frequency of the analysing signal does not correspond to the FFT basis functions, the signal has been windowed before applying the FFT, the note was played using a vibrato effect, or the signal contains inharmonicities or frequency modulations.

In Fig. 4, it can be appreciated that the modified IIR comb filter (which is depicted using a solid line) nulls coincide with the frequencies of the harmonic peaks without changing the amplitude of the E4 harmonics. However, because the sharpness of the filter notches, a considerable amount of energy that

belongs to the harmonic is not extracted, which can lead to inaccurate detections.

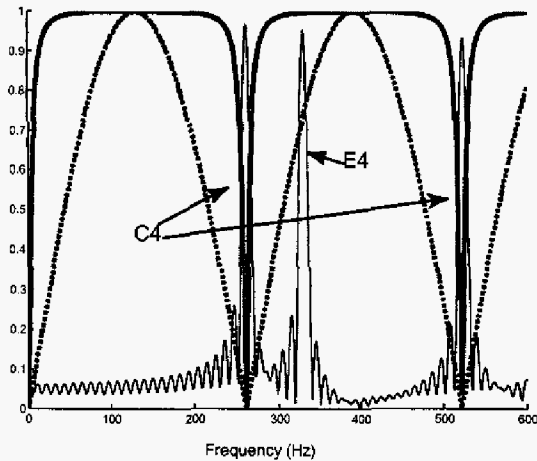


Fig. 4. FIR and modified IIR freq. response

A simple solution would be to modify the parameter ρ , which as becomes closer to 1, the filter notches get sharper. However, by relaxing excessive ρ the filter approximates to the FIR comb filter frequency response.

Ideally, we would like to preserve the filter notch sharpness of the IIR modified comb filter in order to accurately isolate the harmonics of the polyphony, a flat pass-band response between filter nulls to do not alter the amplitude of the unfiltered harmonics, and a relatively wide filter null for the purpose of extracting the entire energy of the harmonic.

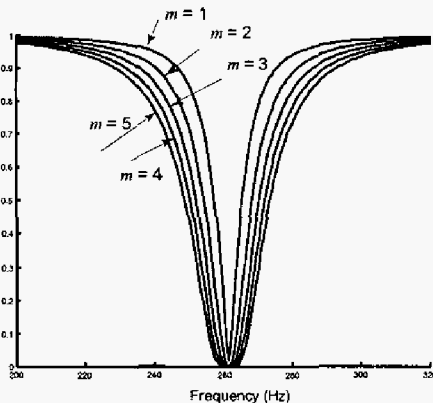


Fig. 5. IIR frequency response depending of m

A good compromise could be obtained by using simple monomials, which order m changes the amplitude of the frequency response as follows:

$$H'(z) = H(z)^m \quad (4)$$

where $H(z)$ is the frequency response of equation 3.

The frequency response of these proposed filters is depicted in Fig. 5, where it can be appreciated that the null gets wider as m is increased, but with the additional cost of a less sharp filter stop-band

4. RESULTS

To evaluate the performance of the algorithm, synthetic signals were utilised to compose a wide range of monophonic and polyphonic signals. The sampling rate used was equal to 44100. The results were compared against the system developed in [8], which is the most relevant multi pitch detector by using comb filters.

The modified IIR comb filters utilised were generated $\rho = 0.999$, and $m = 4$.

4.1. Test 1: monophonic signals

The test material consisted on a database of the 12 semitones of the fourth octave. The results for both systems are illustrated in the first row of Table 1, where it can be appreciated that both systems behave very well, detecting correctly all the notes.

Table 1. Test 1 and 2 results

TEST	Proposed IIR sys.	Existing FIR
1	12/12 = 100 %	12/12 = 100 %
2	22/22 = 100 %	20/22 = 90.9 %

4.2. Test 2: two note polyphonic signals

All the realisable two note mixes derived from the note C were analysed, which are following: [CC#, CD, CD#, CE, CF, CF#, CG, CG#, CA, CA#, CB]

The results can be appreciated in the second row of Table 1. The proposed system detected all the notes. However, the FIR approach missed two notes, detecting incorrectly CD and CG as C#D and GG# respectively.

4.3. Test 3: three note chords

The detection of four common 3 note chords was investigated in this test. The proposed system again detected correctly all notes present in the polyphony. The FIR system missed two notes, detecting C,E,G and C,D#,G as C,C#,G and C,G#,G respectively.

Table 2. 3 note chord detection results

TEST 3	Proposed system	Existing FIR
C,E,G	3/3 = 100 %	2/3 = 66.6 %
C,D#,G	3/3 = 100 %	2/3 = 66.6 %
C,E,G#	3/3 = 100 %	3/3 = 100 %
C,D#,F#	3/3 = 100 %	3/3 = 100 %
Total	12/12 = 100 %	10/12 = 83.3 %

4.4. Test 4: four note chords

In the rich musical context of four note chords, the proposed system detected correctly 3 of 4 chords. However, the system missed 2 notes in one chord, transcribing C,D#,F#, A as C,D#,G, A#. However, this result improves the FIR system, which only detects correctly one chord, transcribing C,E,G, A#, C,D#,F#, A and C,D#,G, A # as C,E,G, F#, C,C#,G, A and C,D#,B, A # respectively.

Table 3. 4 note chord detection results

TEST 4	Proposed sys.	Existing FIR
C,E,G, A#	4/4 = 100 %	3/4 = 75 %
C,E,G, B	4/4 = 100 %	4/4 = 100 %
C,D#,F#, A	2/4 = 50 %	2/4 = 50 %
C,D#,G, A #	4/4 = 100 %	3/4 = 75 %
Total	14/16 = 87.5%	12/16 = 75 %

5. CONCLUSIONS AND FURTHER WORK

In this paper, a multi-pitch estimator is presented. The system utilises modified IIR comb filters, providing better pitch detection results than the most relevant existing FIR comb filter method. This improvement was obtained by using IIR filters, which have a good compromise between the notch sharpness, the pass-band response flatness and the width of the filter null. The results were very encouraging, since the system was capable of detection all the notes in 3 note chords, and 87.5 % in 4 note chords.

However, the system is computationally very intense. In order to improve the speed, polyphase filters can be considered as an area of future work. Also, unexpected frequency components that occur during the onset part of a musical signal, can affect the detection of the multi pitch estimator. Thus, by using an onset detector, the MPE could be notified whether the part of the signal that it being analyzed contains an onset. Thus, the multi pitch estimations will only arise during the harmonic part of the signal.

REFERENCES

- [1] Noll, A.M., *Short -Time Spectrum and "Cepstrum" Techniques for Vocal-Pitch Detection*. 1964. 36: p. 296-302.
- [2] Zolzer, U., *Chapter 9 - Spectral Processing*, in *DAFX - Digital Audio Effects*. 2002, John Wiley & Sons.
- [3] Moorer, J.A., *The optimum comb method of pitch period analysis of continuous digitized speech*. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1974. 22(5): p. 330- 338.
- [4] Brown, J.C., *Determination of the meter of musical scores by autocorrelation*. Journal of the Acoustical Society of America, 1993. 4(94): p. 1953-1957.
- [5] Tolonen, T. and M. Karjalainen, *A computationally efficient multipitch analysis model*. I IEEE Trans. Speech Audio Processing, 2000. 8(6): p. 708-716.
- [6] Abdallah, S.A. and P.M. D. *An Independent Component Analysis approach to Automatic Music Transcription*. in *114th AES Convention*. 2003. Amsterdam.
- [7] Miwa, T., Y. Tadakoro, and T. Saito, *The Problems of Transcription using Comb Filters for Musical Instrument Sounds and Their Solutions*. 2000, TECHNICAL REPORT OF IEICE(written in Japanese)
- [8] Tadakoro, Y., T. Morita, and M. Yamaguchi. *Pitch detection of musical sounds noticing minimum output of parallel connected comb filters*. in *TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region*. 2003.
- [9] Gainza, M., B. Lawlor, and E. Coyle. *Harmonic Sound Source Separation using FIR Comb Filters*. in *117th AES Convention*. 2004. San Francisco.
- [10] Malah, D. and R. Cox. *A generalized comb filtering technique for speech enhancement*. in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '82*. 1982.
- [11] Perlmutter, Y., et al. *Evaluation of a speech enhancement system*. in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '77*. 1977.
- [12] Dattoro, D., *Effect design, part 2: Delay-line modulation and chorus*. Journal of the Audio engineering Society, 1997. 45(10): p. 764-788.
- [13] Soo-Chang, P. and T. Chien-Cheng. *A comb filter design using fractional-sample delay*. in *IEEE International Symposium on Circuits and Systems, ISCAS '97*. 1997.
- [14] Moorer, J.A., *On the segmentation and analysis of continuous musical sound by digital computer*. 1975, Stanford University: CA USA.
- [15] Roads, C., *The Computer Music Tutorial*. 1996: The MIT Press. 904.