Title	Recognition of piano notes(Accepted version)
Author(s)	Foo, Say Wei; Wong, Pui Leng
Citation	Foo, S. W., & Wong, P. L. (1999). Recognition of piano notes. In Proceedings of the International Conference on Information, Communications and Signal Processing 1999: National University of Singapore.
Date	1999
URL	http://hdl.handle.net/10220/4669
Rights	© 1999 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder. http://www.ieee.org/portal/site

RECOGNITION OF PIANO NOTES

Say Wei FOO and Pui Leng WONG

Department of Electrical Engineering
National University of Singapore
4 Engineering Drive 3
Singapore 117576

Abstract

In this paper, an algorithm for the recognition of musical sound from piano is proposed. The music signal is first digitally sampled and transformed from the time domain to the frequency domain using Constant Q Transform. Template matching and other processing techniques are subsequently applied to detect and identify the notes. The algorithm is applied to electronically synthesised music and musical pieces performed on a piano. Perfect recognition is achieved for the electronically synthesised pieces while a high degree of accuracy is achieved for the piano pieces.

1. Introduction

The recognition of music has been an area of interest to many. Much effort has been invested in the identification of sound sources, such as the type of instruments played, and the automatic transcription of musical pieces. In recent years, with the proliferation of personal computers and multimedia systems, research in these areas have gained increasing interest.

In this paper, the pitch tracking aspect of musical recognition is explored. Investigation is carried out on piano notes. A recognition algorithm is developed and tested using both synthesised and real musical pieces.

2. Equal Tempered Scale Representation

Musical notes are periodic complex tones, composed of a fundamental frequency and harmonics [1]. The recognition of a note is thus based primarily on the identification of this fundamental frequency. The Equal Tempered scale representation [2] of musical notes is used in this study. This western musical scale separates adjacent notes by equal intervals of $2^{1/12}$ as depicted in the equation below.

$$f_{k+1} = 2^{(1/12)} f_k \qquad (1)$$

This spacing between adjacent notes is known as one semitone. A semitone spacing of notes corresponds to a 6% frequency separation.

3. Constant Q Transform (CQT)

3.1 Theory of Constant Q Transform

The constant Q spectral transform is developed by Professor Judith C. Brown [4]. It is based on the short time Discrete Fourier Transform [5], but a constant Q value is used instead. Q is defined as the ratio of the frequency *f* over the frequency resolution *df* desired. Mathematically, Q=*f /df*. For semitone spacing, the resolution is $[(2^{1/12})-1]$ =0.059 of the frequency *f*, thus Q=1/0.059 = 17.

In order to constantly resolve at least semitones despite the wide range of frequency values of piano notes ranging from 92.5Hz (F#2) to 4186Hz (C8), the resolution has to be geometrically related to the frequency. For notes at the lower end of the piano keyboard, the resolution must be low while at the high end, it may be higher. Hence, in the constant Q transform, frequencies that are exponentially spaced are sampled. The resolution is now variable and dependent on the frequency.

To calculate the constant Q transform efficiently, Equations (2a) and (2b) are used.

$$X^{cq}[k_{cq}] = \frac{1}{N}\sum_{k=0}^{N-1} X[k]K^*[k, k_{cq}] \qquad (2a)$$

$X^{cq}[k_{cq}]$ is the k_{cq} component of the constant Q transform and X[k] is the Discrete Fourier Transform of a data vector, x[n]. $K^*[k, k_{cq}]$ is the conjugate of the spectral kernels of the transformation given by

$$K[k,k_{cq}] = \sum_{n=0}^{N-1} w[n-(\frac{N}{2}-\frac{N(k_{cq})}{2}),k_{cq}]e^{j\omega k_{cq}(n-N/2)}e^{-j2\pi kn/N} \qquad (2b)$$

$w[n, k_{cq}]$ is a Hamming window that is symmetric about the centre of the interval and *N* is the data frame size. The index *k* corresponds to the indexing of the data vector, while k_{cq} indexes the frequency components in the transform. Details of the derivation of the above equations may be found in [6].

3.2 Implementation of Constant Q Transform

It can be shown that at a sampling rate of S sample per second, the number of samples N required to resolve the minimum frequency f_{min} is related to the Q factor by the following equation:

$$N = S*Q/f_{min} \quad (3)$$

For musical signal sampled at S=22050 sample per second and segmented into frames of N=4096 samples (corresponding to a time interval of about 0.2 second), when Q is set to 17, the lowest frequency 92.5Hz (corresponding to the note F2#) may be identified. With Q=34, the lowest note that may be identified is F3# (185 Hz).

For the proposed algorithm, two frequency ranges of analysis are used. The first range spans from 92.5Hz (corresponding to the note F2#) to 185Hz (the note F3#). A semitone spacing is chosen for this range and Q is set to be 17. This value of Q is chosen to enable the detection of the lowest minimum frequency possible, in this case, 92.5Hz.

The second frequency range starts with the note G3 of frequency 196Hz, and ends with the upper frequency limit of the piano range, C8 (frequency = 4186Hz). As the frequencies of notes in this range are higher, quarter-tone spacing is used to give higher resolution and Q=34.

Note that since two frequency ranges with different Q values are used, two different kernels must be generated using Equation (2b).

The recognition process starts with the extraction of onset of notes from the music piece. The music data is parsed by a time extraction algorithm [9], which determines the instant in time when notes are played. A frame of data of 4096 samples is taken and two Q transforms are performed. Note that two resultant $X^{cq}[k_{cq}]$ vectors corresponding to Q = 17 and Q = 34 are obtained for each frame of music data. To obtain a statistical average of the analysis, several frames of 4096 samples are taken 1024 samples from the start of the previous frame and analysed. This means 3072 samples overlap between successive frames. The number of frames taken depends on the duration between successive notes.

4. Spectra of Piano Notes

When the Constant Q Transform spectral components are plotted on the log-linear graph, the spectral components of a note have fixed positions relative to each other and are independent of the fundamental frequency. The absolute magnitudes of the amplitudes of notes played depend very much on the strength of attack. A strong attack will result in a large value, while weak attacks (where keys are struck softly) will result in small amplitudes [8]. For recognition, we are more concerned with the relative amplitudes of components with respect to one another.

The typical spectra of notes from C4 to G4 are shown in Fig.1. Musical notes do not have the same spectral patterns throughout the entire frequency range. Despite similar decaying patterns, obvious differences can be seen in the spectra of notes. The discerning features between notes manifest themselves in the relative amplitudes of the fundamental and the harmonics. The notes C4 (262 Hz) and G4 (392Hz), for example, have different spectral patterns. However, it is found that notes close to each other have similar spectral patterns. For example, the notes D4, E4 and F4 have similar patterns as C4.

Since notes close to each other share similar spectral patterns, the musical frequencies of interest can be grouped together based on their similarity. From intensive analysis of the spectral patterns of piano notes, three different groups are identified. The first group consists of notes with frequencies starting from 92.5Hz to 190Hz, corresponding

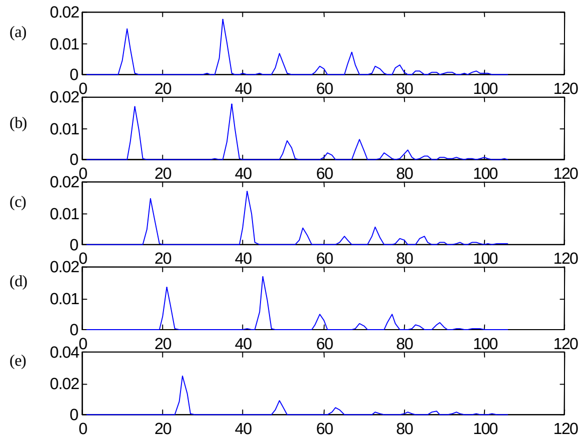

Fig. 1: Spectra of notes (a) C4,(b) D4,(c) E4, (d) F4 and (e) G4

to F2# to F3#. The second group contains notes from 196Hz (G3) to 380.9Hz (F4#), and the third group contains notes from G4 (392Hz) up till C8 (4186Hz).

For the proposed recognition model, it is assumed that notes of the same group have similar spectral pattern and one template may be used for all the notes in the group. Each template is represented by a vector with five frequency components: the fundamental frequency component and the first to the fourth harmonic components. The values of the components of the templates are given in Table 1.

From spectral analysis, it is also established that the principle of superposition holds to a large extent. In other words, the spectrum of notes played together is close to the sum of spectra of the notes played individually. This result agrees with the investigation carried out by Palmer and Brown [8].

Table 1: Component values of the templates

	Template values, $T[k_i]$, where k_i correspond to the k_{cq} value of the i^{th} frequency component				
Freq. group	$T[k_0]$	$T[k_1]$	$T[k_2]$	$T[k_3]$	$T[k_4]$
F2# to F3# (92.5Hz to 185Hz)	0.3750	0.3125	0.1875	0.0469	0.0781
G3 to F4# (196Hz to 380Hz)	0.3080	0.3390	0.1200	0.0900	0.1430
G4 to C8 (392Hz to 4186Hz)	0.5600	0.1990	0.1260	0.0575	0.0575

5. Recognition Algorithm

The proposed algorithm consists of the following steps. First, dot products of the template(s) and the Constant Q Transform spectral components of the frame of data are computed. The values are compared with a predetermined threshold to make a first assessment of the note(s) present. Further processing and statistical analysis are then carried out to improve the accuracy of recognition. These steps are described in detail in the following paragraphs.

5.1 Template Matching

It is expected that the dot product of the components of the matching template and the spectral components of the note when present will yield a higher value than the product of the same set of spectral components when the note is absent. This fact is the basis of the proposed method of recognition.

The dot product between the template vector and the data vector is computed in accordance with the equation below.

$$P_{k_{cq}}[k_{cq}] = \sum_{\substack{\text{corr.}\\ \text{range}\\ \text{of } k_{cq}}} \sum_{k} T_i[k] X^{cq}[k + k_{cq}] \qquad (4)$$

where $k = k_0, k_1, k_2, k_3, k_4$ are five specific component values of CQT with k_0 as the component of the fundamental frequency and T_i is the template for the i^{th} frequency group.

The template $T_i[k]$ is shifted across the Q transformed frame of data and the value of the dot product P_{kcq} is calculated for each value of k_{cq}^{th} component. Each note is represented uniquely by its own k_{cq} value (or values, depending on its tone spacing). By computing the dot product at each k_{cq}, the template is matched with each note to produce a value, which can then be compared with some threshold. This is equivalent to testing if a note is present at the particular frequency. If a note is present, the corresponding spectral components in the data vector, $X[k_{cq}]$ will be high. Thus, when matched with the appropriate template, the resultant dot product, or equivalently, the P_{kcq} value at the particular k_{cq},, will be high.

The template used depends on the range of notes currently being analysed. Template T_1 is used for the group from F2# to F3#, Template T_2 is used for the second frequency group from G3 to F4# and Template T_3 is used for the last frequency group from G4 to C8.

The resultant vector $P_{kcq}[k_{cq}]$ contains the dot products for all the possible notes. The magnitude of P_{kcq} for a particular k_{cq} gives an indication of whether or not the note, with a frequency corresponding to this k_{cq} value, is present.

5.2 Comparison with Threshold

Having obtained the dot product values, the next step is to determine the cutoff value or threshold. If the P_{kcq} value at a particular k_{cq} is above the threshold, it is an indication that the corresponding note is present.

For the proposed model, the threshold values are chosen as a fraction of the maximum value of $P_{kcq}[k_{cq}]$ for a frame. This is to take into consideration the variation of the magnitudes of $P_{kcq}[k_{cq}]$ from frame to frame. Also, different thresholds are determined for different ranges of notes. The threshold values obtained from intensive experiments are given in Table 2.

Table 2: Threshold values for different ranges of notes

Range of notes	Threshold, t
F2# to F3#	0.68
G 3 to B 3	0.20
C 4 to F4#	0.40
G 4 to B 4	0.50
C 5 to C 8	0.90

5.3 Resolving Ambiguity

The magnitude of the first harmonic of a note is comparable with and may even be larger than the magnitude of the fundamental frequency. This may result in ambiguity in identification of notes that are exactly one octave apart. Further processing is required to resolve this ambiguity.

In cases when two notes one octave apart are identified, i.e.,the value of P_{kcq} at a particular note (let's call it P_1) and the value of P_{kcq} at a note one octave higher (let's call it P_2) both exceed the threshold, the following procedure is carried out to determine if indeed there are two notes.

First the ratio P_1/P_2 is determined. The ratio is then compared with the upper limit and lower limit of the confirmation ratios shown in Table 3. If P_1/P_2 is larger than the upper limit, then it is confirmed that the lower note is present but the higher note is absent. If P_1/P_2 is lower than the lower limit, then it is confirmed that the lower note is absent and the higher note is present. If P_1/P_2 falls in between the two limits, then it may be taken that both notes are present.

The values in Table 3 may be obtained from experiments on a large number of piano pieces.

Table 3: Confirmation ratios

Frequency Group	Upper Limit	Lower Limit
F2# to F3#	1.300	0.85
G3 to B3	1.149	0.70
C4 to F4#	1.050	0.85
G4 to C8	1.100	0.85

5.4 Statistical Analysis

By using successive overlap frames of the same note(s), statistical analysis may be performed to improve the accuracy of recognition. It is observed that most of the false notes only appear for some frames during each time instant. The actual notes, on the other hand, either appear continuously or have a higher probability of occurrence. Based on these observations, the following rules may be used to eliminate false notes.

a) Notes that appear continuously for a specific number of frames are considered present.

b) Notes that do not appear continuously for a specific number of frames are considered false notes.

The specific number may be decided through observation and trial and error. Some indicative values are given in Table 4 for two different time intervals between notes.

Table 4: Cutoff numbers

Freq Group	For Time interval ≈ 0.6s, i.e, 10 frames per time instant	For Time interval ≈ 0.3s, i.e, 6 frames per time instant
F2# to F3#	6	3
G3 to F4#	3	2
G4 to C8	4	3

6. Performance Tests

The proposed recognition algorithm is tested with both electronically synthesised piano pieces generated through computer sound card and with piano pieces played by professional on piano.

6.1 Test on Electronically Synthesised Pieces

For all electronically synthesised pieces consisting of both single notes and multiple notes, the recognition rate is perfect. This is not surprising as the synthesised notes have consistent spectra.

6.2 Test on Real Piano Pieces

For the 'real piano pieces', two pieces of music are tested. The musical pieces are obtained from commercially available audio tapes and recorded through a microphone into the computer.

The first piece of real musical piece used is “Menuet”. This is a moderately paced piece. The minimum duration of all the notes in this piece of music is approximately 0.2 second, which meets the minimum frame size restriction for the Q transform.

A high degree of accuracy (98.5%) is attained for this piece of music. Details of the results are summarised in Table 5. Both the three-member chords and all 20 of the two-member chords are resolved perfectly, and 43 out of the total 44 single notes are recognised. The exception is an omission of the note G4 that is played much quicker than 0.6 second as according to the score.

Table 5:Details of test results for "Menuet"

Type of note structure	Present	Recognised	(Type of Error, if any)
Three-member chords	2	2	NIL
Two-member chords	20	20	NIL
Single notes	44	43	Omission

The second piece of piano music, the “Round the Village” is a livelier piece of music compared to “Menuet” and has a faster pace. The recognition accuracy is approximately 87%. Details of the results are presented in Table 6. Upon investigation of the notes that are omitted, it is found that the notes are played too quickly than expected. Whereas for notes that are wrongly recognised, the harmonic contents of the notes played are different from expected.

Table 6:Details of test results for "Round the Village"

Type of note structure	Present	Recognised	(Type of Error, if any)
Three-member chords	1	1	NIL
Two-member chords	24	17	Omission, False Notes
Single notes	40	39	Omission, False Notes

6.3 Variation in Play

The results reveal that although the real pieces are played by professional, there are still some deviations in timing, duration and strength of notes played. These may not be easily detectable by the ears but they affect the accuracy of the proposed recognition algorithm.

7. Conclusion

A method for the automatic recognition of piano notes is proposed in this paper. The method makes use of the Constant Q transform to transform the time domain signals into the frequency domain with constant proportional resolution. Template Matching and additional processing are then applied to detect and identify the note(s).

The algorithm is tested using electronically synthesised pieces and real piano pieces. For the synthesised pieces tested, perfect recognition is achieved. And for the real piano pieces, high accuracy is achieved. Better accuracy may be achieved if the notes are played strictly according to the score. Large variation in play, although not easily detectable by the ears, has an adverse effect on the accuracy of recognition. The algorithm either fails to detect the note(s) or identify the note(s) wrongly.

References

[1] Hartmann, William M., “Pitch, Periodicity, and Auditory Organization”, J. Acoust. Soc. Am. 100, 3491 – 3502, 1996

[2] Sound and Music, University of Cambridge, Study Notes for Option S, 1992

[3] Davidson, Nathan, “Wave RIFF Format and Data Description”, available at internet website: http://www.geocities.com/SiliconValley/Pines/4223/1996

[4] Brown, Judith C., “Calculation of a Constant Q Spectral Transform”, J. Acoust. Soc. Am. 89, 425–434

[5] Peevers, Alan Wesley, “A Real Time 3D Signal Analysis/Synthesis Tool Based on the Short Time, Fourier Transform”, University of California, Berkeley, available at the following internat website: http://cnmat.cnmat.berkeley.edu/~alan/MS-html/msv2.html

[6] Brown, Judith C., “An Efficient Algorithm for the Calculation of a Constant Q Transform”, J. Acoust. Soc. Am. 95, 2698 – 2701

[7] Oppenheim, Alan V., Schafer, Ronald W., “Digital Signal Processing”, Prentice Hall of India Private Limited, Chapter 3, 1996

[8] Palmer, Caroline, Brown, Judith C., “Investigations in the Amplitude of Sounded Piano Notes”, J. Acoust. Soc. Am. 90, 60 - 65

[9] Yew-Choe, Lum,, “Detection of Onset of Musical Notes", 1999 Final Year Project Report, Dept of Electrical Engineering, NUS.