ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 14, No. 3, 2019, *pp*.178-183

Recognition of piano keyboards based on sound feature extraction

Lun Li [+], Cheng Li [*]

School of Computing, University of Kent, CT2 7NZ, England, UK
(Received May 10 2019, accepted June 11 2019)

Abstract. The tones of piano are produced by regular periodic oscillations and different keys' individual inherent audio characteristics. Based on the collection of the sound wave signals and the analysis of the frequency characteristics from different keys, this study explores the methods to effectively identify and detect the corresponding numbers of keys. This paper will also compare and analyzed the different methods of audio signal extraction and identification of piano keys, to find effective methods to make theoretical preparation for further system development.

Keywords: Piano key identification, sound feature extraction, Pattern Recognition, Numerical analysis

1 Introduction

Piano is the most representative member of musical instrument family because of wide range, bright timbre and varied tones. Numerous of piano work has enriched human culture throughout the history of music. Social development promotes the piano learning and practice into daily life of many families. However, there is a significant difference between the learning of general knowledge and learning piano. Learning piano requires on-site guidance and repeated practice by experiences instructors [1,2].

Essentially, every piano piece is an audio collection composed of different keys, that produce different frequency characteristics. Therefore, it is expected to effectively identify and detect the corresponding keys' serial numbers by analyzing the different audio features of the keys. And it is quite helpful for the guidance and specification of the piano practice process, which mainly includes sound wave feature extraction and piano key recognition. Tone is a property of frequency signal which is not relative to the contextual content. Different tones signal analysis parameters can be divided into time domain, frequency domain, cepstrum domain and so on. The Mel Frequency Cepstrum Coefficient (MFCC) has high recognition performance and anti-noise ability. The recognition of musical signal can be achieved through the characteristics of musical signal extracted by pre-processing of musical signal. Currently, the methods of acoustic signal recognition mainly include Dynamic Time Warping algorithm (DTW), Vector Quantization (VQ) and Hidden Markov Model (HMM) hybrid technology [3~5]. This paper aims to compare and analysis piano audio signal extraction and recognition methods, in order to develop effective methods to make theoretical preparation for the further development of system software. Therefore, develop a scientific tool which enable piano leaner to be less dependent on piano instructors.

2 Extraction and analysis of tones signals

Piano's tone is created by regular periodic vibrations of the strings， which has pitch, time value, loudness, timbre four basic characteristics and corresponding to the frequency, duration, amplitude and spectrum distribution of vibration. The standard piano has 88 keys (36 black keys, 52 white keys) corresponding to 88 strings respectively. By pressing the keys and striking the strings. And order the frequency from low to high(left to right),denoted as $X_i(i = 1, \cdots, 88)$. The fundamental frequency goes from 27.5Hz, 29.12Hz to 4786Hz. The 29th white key is the international standard tone, and its fundamental frequency is 440Hz. Based on this fundamental frequency, we establish the basic model library of audio recognition.

2.1 Preprocessing of tones signals

A continuous piece of music is composed of many single-tones in chronological order, including different fundamental frequency, it is a typical time-varying signal. The frequency domain composition of a single note

[+] Corresponding author. Tel.: 0044-7561589118; *E-mail address*：ll375@kent.ac.uk
[*] Working at HiSilicon Technologies (HuaWei)

is stable, and the fundamental tone and overtone are completely fixed, frequency is constant, only the amplitude reduces over time. The collected continuous audio signal sequence needs to be divided into single note signals by denoising, pre-emphasis, end-point algorithm and so on.

Sound signals with low signal-to-noise ratio should be separated from music and noise. Blind source separation, wavelet threshold denoising technique, separation methods based on statistical methods all have different denoising results.

After the denoising, it needs to pre-emphasis the tones signals. The function of pre-emphasis is to amplify the high-frequency of sound by increasing the high-frequency part of the musical signal， which made the frequency spectrum of tones signal flattened for spectrum analysis and it is realized by first-order Digital filter.

The non-stable tones signal is transformed into a short-term stationary signal. The use of overlapping segmentation can be used to maintain the continuity of smooth transition between two frames. Take the frame length as 10～30ms, the overlapping part of two frames is called frame shift, and the ration of frame shift to frame length is 0～1/2.Framing can be weighting through moveable finite length window, which is the product of window function $w(t)$ and sound signals $x(t)$ to build

$$x_w(t) = w(t)x(t) \quad (1)$$

The most common use of window functions are rectangular window, hamming window, hanning window and blackman window.

The use of end-point detection algorithm can accurately detect the starting-point and the end-point of tones signals, which can realize the divide of single notes. The double threshold algorithm based on short-term energy and zero-crossing rate is the common algorithm of endpoint detection. Before the endpoint detection, two thresholds are set for the short-term energy and zero crossing rate. If the signal exceeds the high threshold with a large margin, it indicates that the signal has reached a certain intensity, which can be mostly determined it is caused by the tones signal.

The endpoint detection of tones signal is divided into mute, transition, musical segment and end. In the mute segment, the starting point is marked if the energy or zero-crossing rate exceeds the low threshold. In the transition part, if the two parameter values fall back to the low threshold at the same time, the current state will be restored to the mute state; otherwise, it can be confirmed that the music tone segment has been entered.

2.2 Time-frequency domain analysis of music signal

Time domain analysis is to analyze the time domain waveform of music signal, Extracting the time domain parameters of music signal. It is generally used for the most basic parameter analysis and can intuitively represent the musical signal. The time domain parameters of music signal include short time energy, short time average amplitude, short time average zero crossing rate, short time auto-correlation function and short time average amplitude difference function.

The frequency domain analysis of musical signal includes spectrum, power spectrum, cepstrum, spectrum envelope analysis, etc. The commonly used methods include band-pass filter bank, Fourier transform, linear prediction and wavelet transform.

Spectrum analysis

Spectrum analysis mainly relies on the short-time Fourier transform method. The short time Fourier of signal $x(n)$ transformed as

$$X_n(e^{j\omega}) = \sum_{m=-\infty}^{+\infty} x(m)\cdot w(n-m)e^{-j\omega m} \quad (2)$$

$w(n)$ is window sequence.

Signal $x(n)$'s cepstrum $c(n)$ is defined as inverse Z transformation, which is

$$c(n) = Z^{-1}[\ln|X(n)|] \quad (3)$$

The commonly used characteristic parameters are linear predictive cepstral coefficients (LPCC) and Mel frequency cepstral coefficients (MFCC). MFCC has good noise resistance and recognition ability, which is used here to extract parameters of musical signal features.

$$c_t(n) = \sum_{k=1}^{m} \log S_t(k)\cos[\pi(k-0.5)\frac{n}{M}], \qquad n = 1,2,\cdots,L \quad (4)$$

MFCC coefficient only reflects the static characteristics of musical signal, and the first order difference of ΔMFCC is generally used to reflect the dynamic characteristics of musical signal.

3 Recognition of tones of piano keys

The musical signal is expressed in the form of characteristic parameters, and the characteristic data with the same characteristics are obtained through corresponding processing to form the reference pattern library, and then the characteristic parameters to be recognized are matched with the reference pattern library to determine the most similar notes.

3.1 Dynamic Time Warping

Dynamic time warping algorithm (DTW) is a nonlinear warping technology combining time warping and distance measurement calculation, which is implemented by dynamic programming technology (DP). Assuming the test musical parameters have i frame vector, reference template has j frame vector. When $i \neq j$, DTW algorithm is to find an optimal time structuring function $j = \omega(i)$, which maps the time axis i of the test vector to the time axis j of the template non-linearly, so as to minimize the cumulative distortion, which is ω function satisfies

$$D = \underset{\omega(i)}{Min} \sum_{i=1}^{l} d[T(i), R(\omega(i))] \tag{5}$$

The following is an example of the "Bullet" style, which you may want to use for lists.The distance measure $d[T(i), R(\omega(i))]$ between the i frame test $T(i)$ vector and the frame j template vector $R(j)$, also it is the distance between the two vectors under the optimal time normalization.

3.2 Vector Quantization Technique

Vector quantization is a data compression and coding technology， the basic principle of it is to extract feature vectors from several(such as frames) music signals, quantize them once in multi-dimensional space, and compress data with less loss of information [7].

Each K number of continuous sample points of musical signal sequence are divided into a group. Assume N there be K-dimensional vector in total

$$X(i) = \{x(1), x(2), \cdots, x(k)\}, \quad i = 1,2,\cdots, N \tag{6}$$

All K-dimensional vector create a Euclidean space R^K, and separate into J number of non-intersecting subspace $R_1, R_2, \cdots, R_J$. $Y = \{Y_j \in R^K, 1 \leq j \leq J\}$is a vector quantizer containing J number of vectors is called code book, and Y_j is called code. The process of vectorization involves the measure of distortion when the boundary of the partition region is compared with the vector. In 1980, LBG (by Linde, Buzo and Gray) algorithm is a classical vectorization clustering algorithm which is designed through vector quantizer code book. Firstly, find the center of a training sequence and generate an initial code book by splitting method. Finally, the training sequence is grouped according to the elements in the code book, find out the center of each group, get a new code book, and then replace the new code book to repeat the above process[6].

3.3 Hidden Markov Model

Hidden Markov model is a statistical model based on Markov chain, which is widely used in voice processing and recognition. It is a double random process, one is the transition of Markov chain description state; The second is the statistical correspondence between the description state of the random process and the observed value. The basic HMM algorithm includes forward-backward algorithm, Viterbi algorithm and Baum-Welch algorithm.

3.3.1 Forward-backward algorithm

Forward algorithm Forward probability variable is defined as

$$\alpha_t(i) = P(o_1, o_1, \cdots, o_t \mid q_t = i, M) \tag{7}$$

So $\alpha_t(i)$ and $P(O|M)$ can be derived from following formula

DOI https://doi.org/2024-JICS-22411 | Generated on 2025-04-05 22:12:26

$$\begin{cases} \alpha_1(i) = \pi_i b_i(o_1), 1 \le i \le N \\ \alpha_t(i) = [\sum_{i=1}^{N} \alpha_t(i) a_{ij}] b_j(o_{t+1}),\ 1 \le t \le T-1, 1 \le j \le N \\ P(O \mid M) = \sum_{i=1}^{N} \alpha_T(i) \end{cases} \quad (8)$$

where $b_j(o_{t+1}) = b_{jk}|_{o_{t+1}} = v_k$.

Backward Algorithm The backward probability variable is defined as

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \cdots, o_T \mid q_t = i, M) \quad (9)$$

where $\beta_t(i)$ and $P(O|M)$ can be derived from following formula

$$\begin{cases} \beta_1(i) = 1 \\ \beta_t(i) = \sum_{i=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), t = T-1, T-2 \cdots, 1,\ 1 \le j \le N \\ P(O \mid M) = \pi_i b_i(o_1) \beta_1(i) \end{cases} \quad (10)$$

3.3.2 Viterbi Algorithm

Defined $\delta_t(i)$ as time t, followed path $q_1, q_2, \cdots, q_t$, and $q_t = i$ and produced $o_1, o_2, \cdots, o_t$,

$$\delta_t(i) = \underset{q_1, q_2, \cdots, q_t}{Max} P(q_1, q_2, \cdots, q_t,\ q_t = i, o_1, o_2, \cdots, o_t \mid M) \quad (11)$$

then

$$\begin{cases} \delta_t(i) = \pi_i b_i(o_1),\ \varphi_t(i) = 0, 1 \le i \le N \\ \delta_t(j) = \underset{1 \le i \le N}{Max}[\delta_{t-1}(i) a_{ij}] b_j(o_t),\ \varphi_t(j) = \arg \underset{1 \le i \le N}{Max}[\delta_T(i) a_{ij}], 1 \le j \le N, 2 \le t \le T \\ P = \underset{1 \le i \le N}{Max}[\delta_T(i)],\ q_T = \arg \underset{1 \le i \le N}{Max}[\delta_T(i)] \\ q_t = \varphi_{t+1}(q_{t+1}),\ t = T-1, T-2, \cdots, 1 \end{cases} \quad (12)$$

3.3.3 Baum-Welch algorithm

From formula (10) and (11) Forward variable and backward variable, have

$$P(O \mid M) = \sum_{i=1}^{N} \sum_{j=1}^{N} a_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j),\ 1 \le t \le T-1 \quad (13)$$

Find out M to maximizes $P(O|M)$ is a functional extremum problem, Because the given training sequence is limited, there is no best way to estimate M. Baum-Welch uses the idea of recursion. Make $P(O|M)$ locally extremely large and get the final mode parameter $M = \{A, B, \pi\}$.

Defined $\xi_t(i,j)$ for a given training model M and training sequences O, at time t of markov chain state i and time t+1 of state $j's$ probability. It follows that

$$\xi_t(i, j) = [a_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)] / P(O \mid M) \quad (14)$$

Then，the probability that the Markov chain is in state i at time t is

$$\gamma_t(i) = P(q_t = i \mid O, M) = \sum_{j=1}^{N} \xi_t(i, j) = a_t(i) \beta_t(i) / P(O \mid M) \quad (15)$$

So, $\sum_{t=1}^{T-1} \gamma_t(i)$ indicate the expected number that transferred out from i, $\sum_{t=1}^{T-1} \xi_t(i,j)$ indicates the expected number transferred from state i to state j. Based on this, the well-known re-estimation formula of the Baum-Welch algorithm is estimated

$$\begin{cases} \bar{\pi}_i = \gamma_1(i) \\ \bar{a}_{ij} = \sum_{t=1}^{T-1}\sum_{j=1}^{N}\xi_t(i,j)/(\sum_{t=1}^{T-1}\gamma_t(i)) \\ \bar{b}_{jk} = \sum_{t=1:o_t=v_k}^{T-1}\gamma_t(i)/\sum_{t=1}^{T-1}\gamma_t(i) \end{cases} \quad (16)$$

Therefore the solving process of HMM parameter model $M=\{A,B,\pi\}$ is according to the observation sequence O and selected initial model $M=\{A,B,\pi\}$, a new set of parameters $\{\bar{\pi},\bar{a}_{ij},\bar{b}_{jk}\}$ is obtained through the calculation of the revaluation formula. So we can get new model $M=\{\bar{A},\bar{B},\bar{\pi}\}$, iteration until $P(O|M)$ converges.

4 Numerical experiment and analysis

4.1 Experiment of single key's tone

Dynamic time warping algorithm (DTW) is a nonlinear warping technology combining time warping and distance measurement calculation, which is implemented by dynamic programming technology (DP). Assuming the test musical parameters have i frame vector, reference template has j frame vector. When $i \neq j$, DTW algorithm is to find an optimal time structuring function $j=\omega(i)$, which maps the time axis i of the test vector to the time axis j of the template non-linearly, so as to minimize the cumulative distortion, which is ω function satisfies

In the experiment, the piano has 88 keys. In a quiet indoor environment, three consecutive notes are collected from each key, which are stored in a sound file. The 88 keys are numbered from 1 to 88 in order from left to right and from low to high frequency. Endpoint detection and note segmentation are performed by double-threshold detection algorithm based on short-term energy and zero-crossing rate, a total of 364 single note file databases can be split. And divide it into 3 groups (k=1,2,3), each set contains a single note of 88 keys and calculate the cepstrum coefficients of Mel frequencies for each note. One group was used as test samples, and the other two groups as training samples were tested as DTW, VQ and HMM, respectively. The results are shown in table 1.

Table 1 Comparison of results of three identification methods for single note file

	1st set	2nd set	3rd set	Recognition Rate（%）
DTW	85	88	86	98.11
VQ	88	88	88	100
HMM	88	88	88	100

As can be seen from table 1, in terms of the recognition accuracy, both VQ algorithm and HMM algorithm achieve 100% accuracy in the database recognition and the recognition of simple piano music. The recognition rate of DTW algorithm is 96.59% and 95.45%. Therefore, the recognition accuracy of VQ algorithm and HMM algorithm is slightly higher than that of DTW algorithm.

4.2 Continuous music experiment

In order to verify the overall recognition effect, the collected piano music comes from the part of "Jingle Bell Rock", with a total of 89 notes. Total of 364 single note file databases can be split. A single note segmentation will perform on a continuously played music sample, draw its time domain waveform, the average short-term energy and the average zero crossing rate, and the beginning and end of each note. Similarly, VQ and HMM algorithms with better recognition rate can be used continuously to compare and match the musical signal characteristics in the database. Both methods can achieve the 100% recognition rate, but the VQ running time is less than HMM（see Tab.2）

DOI https://doi.org/2024-JICS-22411 | Generated on 2025-04-05 22:12:26

Table 2 Comparison of results of three identification methods for Continuous music

	DTW	HMM	VQ
Recognition Numbers	86	89	89
Recognition Rat（%）	96.6	100	100
CPU time (s)	53.5	98.2	71.6

Based on the analysis of the Recognition Rat and CPU times, VQ algorithm has a high recognition accuracy and a fast recognition speed. In this paper, we evaluated that the combination of note Mel frequency ceptrum coefficient and VQ algorithm, and the results suggest that combining them can have a high piano key recognition rate.

References

[1] Delfs C, Jondral F. Classification of Piano Sounds Using Time-Frequency Signal Analysis.

[2] IEEE International Conference on Acoustics, 2002 ,Vol. 3 No3,pp2093

[3] Anders Thorin，Xavier Boutillon，José Lozada. Modelling the dynamics of the piano action: is apparent success real? Acta Acustica united with Acustica, 2014, Vol 100, No. 6,pp. 1162-1171

[4] Davis, S. Mermelstein, P. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences.1980, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 28 No. 4, pp357-366

[5] Gernot A. Fink. Markov Models for Pattern Recognition. 2008 Springer-Verlag Berlin Heidelberg

[6] Viterbi A.J. A personal history of the Viterbi algorithm. IEEE Signal Processing Magazine, Vol. 23 No. 4, pp120 - 142

[7] Wilpon J G，Rabiner L R，Lee C L. Auto recognition of keywords in unconstrainte speech using hidden markov models. IEEE Trans. ASSP，1990, Vol. 38 No.11：pp1870-1878.

[8] Linde Y，Buzo A，Gray R M. An algorithm for vector quantizer design. IEEE Trans. Communication，1980,2, 8 pp84-95.

[9] Stephane Mallat， Sifen Zhong. Characterzation of Signals from Multiple Edges. IEEE Transactions on Pattern Analysis and Machine Intelligence， 1992,Vol.14 No.7.

[10] Yasushi Horii, Tsutumi M.. Comparison between musical acoustics parameters of an upright and a grand piano. Journal of the Acoustical Society of America, 2003, No.10

[11] Plitsis G . Sound feature extraction to distinguish between a grand and an upright piano. Journal of the Acoustical Society of America , 2008 , Vol.123 No.5, pp3801