

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 learner 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 $X_i(i = 1, \dots, 88)$. 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.

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

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