Report
Automatic Polyphonic Piano Music Transcription

Agnieszka Szefer

June 5, 2014

Todo list

Contents

1 Introduction

This chapter outlines the motivation behind the project, its main objectives and contributions. At the end of the section we describe a general outline of the report.

1.1 Motivation

Even after millions of years of evolution of our sense of hearing it is still very difficult for a human without any prior musical education to tackle the problem of transcribing music, i.e. writing down the notes that make up a given piece of music. Furthermore, transcribing complex music pieces is a time-consuming task. Even for naturally talented people it may take several attempts to transcribe a whole piece correctly.

Automating the process of music transcription could be therefore of great help to musicians, especially to those lacking the skill of transcribing a piece 'by ear'. There are large amounts of music that are not available in an annotated form. This includes traditional music passed from generation to generation or just released popular songs. Furthermore, there is an increasing community of people interested in learning how to play a piece from sheet music. One of the most interesting applications of automated transcription is real time or offline feedback for music students. A student could see if what he or she plays is correct in respect to the sheet music he or she was given.

Reference to Music Prodigy

Why piano music

1.2 Objectives

The aim of this project was to create an end-to-end system for automated transcription of piano music. In particular, we focused on correct detection of polyphonic music, it.e. where more than one note can be played at a time. We also investigated current state of the art methods for pitch detection and looked for ways to improve them in our system.

Improve the last sentence.

1.3 Contributions

In this report we complement the state of the art in automatic music transcription with the following contributions:

- We present an end-to-end solution to automatic music transcription.

  Can this be a contribution?

  with user inferface maybe?

- We introduce two techniques for finding noise threshold (one using least squares approach and another the total spectrum of significant spectral peaks) that improve the detection of pitch candidates.
- We filter out insignificant spectral peaks in an early stage of transcription what improves the speed of pitch detection.

  Does it count as a separate contribution?

- We show that using one FFT instead of commonly used STFT for each sound frame is 'good enough' for pitch detection in our system and also requires less computations.
- We experiment and test using real life data recorded by the author, where most of research papers test on studio recorded data.
- We use user input data for better noise estimation and more accurate pitch detection.

  Is this a separate contribution?

- We present challenges encountered in the development of the system and explain how we overcame them.
- We evaluate our system and show its limitations, and suggest ideas for future work that could improve current transcription solutions.

1.4 Report Outline

Describe report outline

2 Background

This chapter introduces some concepts necessary to understand this project. We provide background in the area of digital signal processing and piano music. We also describe the state of the art in the field of automated music transcription.

2.1 Sound

Musical instruments or any other sources generate vibrations that are propagated through air or other medium in a form of a waveform. These vibrations are commonly known as *sounds* and we are able to hear them because of the changes in the air pressure in our ears. If the pressure varies according to a repeating pattern we say that the sound has a *periodic waveform* [Roads, 1996].

One such pattern repetition is called a *cycle*, and the *fundamental frequency* is the number of cycles that occur per second. The frequency is measured in units of Hertz, where 1Hz means one cycle per second. The *period* of a waveform is the lenght of the cycle and as it increases the frequency decreases.

Humans are usually able to hear frequencies in the range between 20Hz and 20000Hz.

2.1.1 Sound Wave Representations

One way of representing a sound waveform is as a time-domain graph showing how the air pressure changes over time (see Figure 1). The *amplitude* is the maximum displacement of the wave measured from its equilibrium position.

A waveform may consist of not only the fundamental frequency, but also many other frequencies. The frequency-domain representation, also called the *spectrum*, shows the frequencies that contribute to the sound (see Figure 1). Any frequency component is a *partial*. A partial that is an integer multiple of the fundamental frequency is called a *harmonic*. If we assume a fundamental frequency of 440Hz, its second harmonic is 880Hz, third harmonic is 1760Hz, etc. Any frequency higher than the fundamental frequency is an *overtone* (see Figure 2).

2.2 Digital Signal

Figure 2.2 illustrates the process of digital audio recording and playback. A source generates sound waves. A microphone transduces the air pressure produced by this source into electrical voltages. The voltages are passed to analog-to-digital-converter (ADC). At each tick of the sample clock the ADC converts the voltages into strings of binary numbers.

2.2.1 Sampling

In contrast to the analog signal (see Figure 3), the digital signal is defined only at the points of time it has been *sampled* at. In Figure 4 each vertical bar represents one sample of the signal. The *sampling frequency* or *sampling rate* is expressed in units of Hertz.

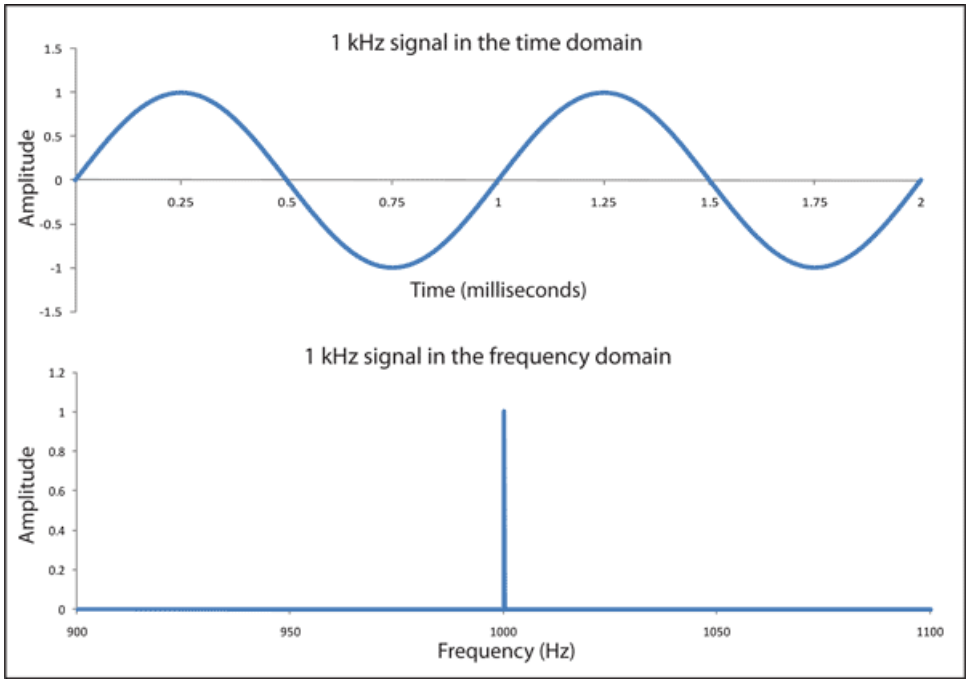


Figure 1: Time-domain and frequency-domain representations of a sound wave.

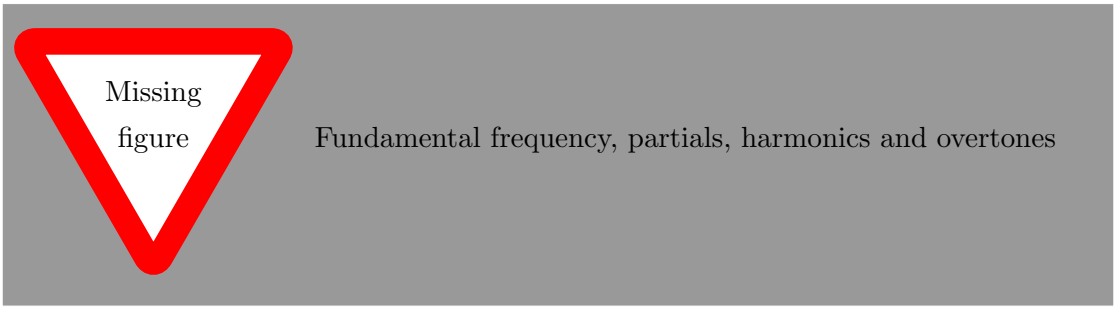


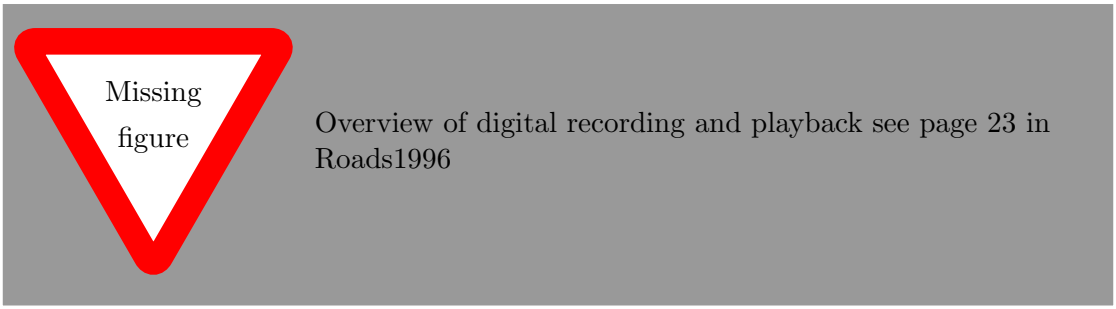
Figure 2: Fundamental frequency, partials, harmonics and overtones.

In this project we experimented with signals sampled at the most common sampling frequency: 44.1 KHz with 16-bit samples.

Mention aliasing and quantization error?

Some files were recorded at 48000Hz

Add labels to the diagrams.



Overview of digital recording and playback see page 23 in Roads1996

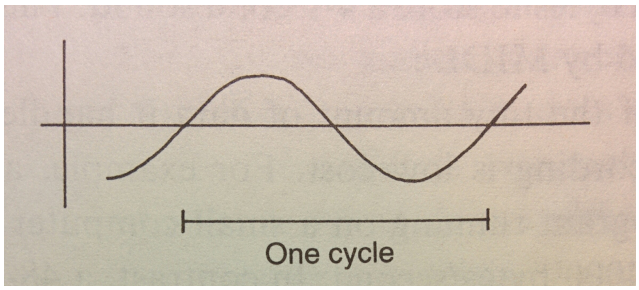


Figure 3: Analog representation of a signal [Roads, 1996].

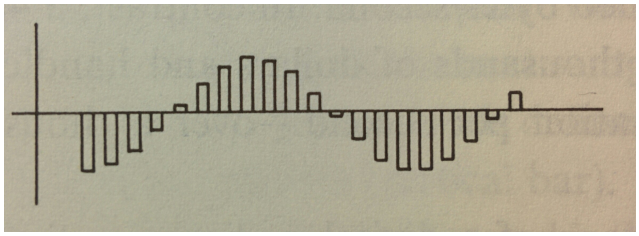


Figure 4: Digital representation of a signal [Roads, 1996].

2.2.2 WAVE Format

It may be better to move it to Outline section and explain why this format is better than others.

WAVE is an audio file format used for storing audio bitstreams. This format is uncompressed what ensures the highest quality of the recorded sound. Moreover, it is the most general music storage format and can be easily converted to other popular ones such as .mp3.

Mention stereo (2 channels) vs. mono sound.

2.3 MIDI

MIDI is a popular protocol for control of digital music systems that allows for communication between an electronic instrument and a computer. In contrast to a digital audio recorder, a MIDI sequencer does not transmit the sampled waveform of the sound. When a music piece is played on a keyboard, a MIDI sequencer records only the start and ending time of each note, its pitch, and the amplitude of the beginning of a note. Therefore, AaStandard MIDI File (SMF) requires much less storage than .WAV to represent similar data.

For instance, if you play 4 quarter notes at a tempo of 60 beats per minute on a MIDI synthesizer, just 16 pieces of information of this 4-second sound are captured (4 starts, ends, pitches and amplitudes). On the other hand, if you record the same sound with a digital audio recorder set to a sampling frequency of 44.1 KHz, 352,800 pieces of information are recorded (44,100 * 2 channels * 4 seconds). Using 16-bit samples, it takes over 700,000 bytes to store a 4-second sound. This is 44,100 times more data than is stored by MIDI [Roads, 1996].

2.4 Piano

Piano is one of the most popular musical instruments. There are two types of pianos: a grand piano (Figure 5(a)) and an upright piano (Figure 5(b)). A piano has usually 88 keys (52 white and 36 black). The lowest note is A0 with a fundamental frequency of 27.5Hz and the highest one is C8 with a fundamental frequency at 4186.0 Hz.

When you strike a piano key a padded hammer hits steel strings. The hammer rebounds and the strings continue to vibrate at their resonant frequency. When you release the key, a damper stops the string's vibration.

(a) A grand piano (b) An upright piano

Figure 5: Types of pianos.

How the sound is made.
Inharmonicity.
Tuning.
Pedals.

Bibliography

Curtis Roads. *The Computer Music Tutorial.* The MIT Press, 1996.