Real-time monophonic guitar pitch estimation
The feasibility of Fourier transform based methods

Luc de Jonckheere

October 5, 2021

Abstract

Short summary.

1 Introduction

Pitch estimation, which is also referred to as f0 estimation, is an important subtask within the field of Automatic Music Transcription (AMT). The goal of pitch estimation is to estimate the pitch or fundamental frequency f_0 of a given signal. In the context of AMT, pitch estimation is used to determine what note is played in a given signal.

Real-time pitch estimation is a subproblem where we want to estimate the note associated with the measured pitch while the musician is playing it with minimal latency. This entails we have to use the latest received signal. In contrast to non-real-time methods, we have no knowledge of what may happen ahead of time and signal corresponding to previous notes is irrelevant. This limits the methods we can use to solve this problem.

If pitch estimation can accurately be performed in real-time, it can be used to create a digital (MIDI) instrument from an acoustic instrument. This digital instrument can then be used as an input for audio synthesizers, allowing musicians to produce sounds from a wide variety of instruments. Furthermore, accurate real-time pitch estimation can be used to automatically correct detuned instruments by pitch shifting the original signal to the closest harmonious note.

Our research focusses on monophonic pitch estimation. Here, we assume that the signal contains at most one note. It is much easier to perform monophonic pitch estimation compared to polyphonic pitch estimation, especially when using Fourier transform based methods, as fundamental limits of the Fourier transform inhibit our ability to discern two low pitched notes [3]. Furthermore, hexaphonic guitar pickups are becoming more widespread, which allows us to view the guitar as six monophonic instruments instead of one six-way polyphonic instrument.

This thesis builds upon a preliminary research project [4]. In our research project, we found that Fourier transform based pitch estimation methods might not be well suited for real-time use due to fundamental limitations of the Fourier transform. In this work, we will further research if Fourier transform based methods are viable, as real-time transcription research often relies Fourier transform based methods.

The goal of this thesis is to research the limits of Fourier transform based real-time pitch estimation. To correctly assess the limits, we develop a pitch estimation framework which is available at www.github.com/lucmans/digistring. This framework will focus on extensibility and the ability to perform automated tests. This is important, as much work in this field does not provide its associated source code. This limits the ability to build on other's work and hinders direct comparisons between different methods. Our framework can provide a common ground for the different methods and algorithms to be implemented and compared in.

2 Related work

Other papers. Sample citations [5] [7] [1].

3 Preliminaries

Jargon required to understand this paper.

3.1 Real-time

What we mean with real-time vs how other work uses real-time. Formal vs common definition.

3.2 Fourier transform

Emphasis on discrete FFT (where most papers only mention continuous, which causes misconceptions).

Effect of sampling rate (which is often chosen too low) and frame size. Sensitivity for specific frequencies. Quadratic interpolation.

3.3 Fundamental, overtones and transients

Lorum Ipsum.

3.4 Music notation/theory

In modern western music, we use the twelve-tone equal temperament (12-TET) music system. This system divides an octave, which is the interval between a pitch and another pitch with double the frequency, into twelve equally spaced semitones on the logarithmic scale. The logarithmic scale is used such that the perceived interval between two adjacent notes is constant [6]. Moreover, the ratio between two frequencies in an n-semitone interval is $\sqrt[12]{2}^n$ or $2^{\frac{n}{12}}$, invariant to pitch.

Using scientific pitch notation, every note can be uniquely identified by combining the traditional note names A to G (with accidentals such as $\sharp$ and $\flat$) with an octave number (e.g. $E_3^\flat$). An octave starts at C, which means the octave number increases between B and C. This counter intuitively implies that A_3 is higher than C_3. Note that in 12-TET, $C_4^\sharp$ and $D_4^\flat$ are enharmonically equivalent.

The 12-TET music system only describes the relation between two notes in an interval. In order to play with other musicians in harmony, an arbitrary note has to be tuned to a specific frequency. Per ISO 16, the standard tuning frequency of the A_4 is 440 Hz within an accuracy of 0.5 Hz [2].

The range of a guitar in standard tuning is from E_2 to E_6... TODO

4 Paper content

Actual research etc.

4.1 Software amplification

(Software representation of samples and FFT). The FFT works on floating point numbers but most audio interfaces give up to 24 bit integers... We found empirically that when amplifying the input signal in software, peaks in the frequency domain are much easier to detect. However, it has to be done carefully to prevent distortion artifacts.

5 Implementation

Details about the program I've written. Usage instructions, code choices, code structure, screenshots, expandability

6 Experiments

Test the system.

7 Conclusions

What we did in this paper. Reflection on the performance of the system. Final reference to the source code/dataset location.

8 Future work

What could still be improved/further researched.

References

[1] FFTW3 page about two times speed-up by omitting complex part. http://www.fftw.org/fftw3_doc/One_002dDimensional-DFTs-of-Real-Data.html Last accessed on 12-04-2021.

[2] Iso 16:1975. acoustics — standard tuning frequency (standard musical pitch). 1975.

[3] Eric J. Anderson. Limitations of short-time fourier transforms in polyphonic pitch recognition. Ph.d. qualifying project report, University of Washington, Department of Computer Science and Engineering, 1997.

[4] Luc de Jonckheere. Real-time guitar transcription using Fourier transform based methods; a pitch estimation framework and overtone sieve algorithm. 2021.

[5] S.S. Limaye K.A. Akant, R. Pande. Accurate monophonic pitch tracking algorithm for QBH and microtone research. *The Pacific Journal of Science and Technology*, 11(2):342–352, 2010.

[6] Michael J. O'Donnell. Digital sound modeling: Perceptual foundations of sound. http://people.cs.uchicago.edu/~odonnell/Scholar/Work_in_progress/Digital_Sound_Modelling/lectnotes/node4.html Last accessed on 04-10-2021.

[7] K. M. M. Prabhu. *Window Functions and Their Applications in Signal Processing.* CRC Press, 2013.