

Music REgenerator

SAMUEL DUPONT

LUCAS BOURNEUF

Samuel.dupont.etu@univ-lemans.fr

November 26, 2016

Last update: December 28, 2016

Abstract

This report is written to explicit the different parts of the music regenerator project. The global aim is to generate non random music using statistical tool which is listenable and sounds good. It's based on the decomposition of one or several musics in an index of probability of the note to play. This decomposition is achieved using Markov chain method.

The music is based on two main parameters which are the pitch of the note (the midi notation is used), and the note position. This result in the construction of two Markov chains one for the rhythm and one for the note. Both of them are for now independent and generate a music which is acceptable to hear let a big margin for improvements.

A part of this project is also dedicated to the transcription of music into MIDI format using pitch recognition algorithm and onset detection. It explores two methods of pitch reconignition one based on a modified cepstrum and another based on a modified autocorrelation method. The results are good for a part of the musical spectra however in the high frequency domain it becomes less effecting resulting in the wrong estimation of the played note.

Chapter 1

Dsp and audio algorithm for pitch recognition

I. INTRODUCTION

In order to be able to convert a musical sample from a wav file for example, it needed to accurately determine the pitch of the different notes. This chapter will present two different methods based on the work of Philip McLeod [1]. There are based on the decomposition of the note into several smaller windows where the algorithm is applied, then the results of all the windows are reunified to achieve the final pitch estimation.

A first section will present the two different algorithms, the second will explain the pre processing steps and then how to assemble the results. Next the evaluation of the methods will be discussed before the conclusion.

The figure 1.1 shows an input signal of 4024 points corresponding to midi 50 played with a piano used for the following figure.

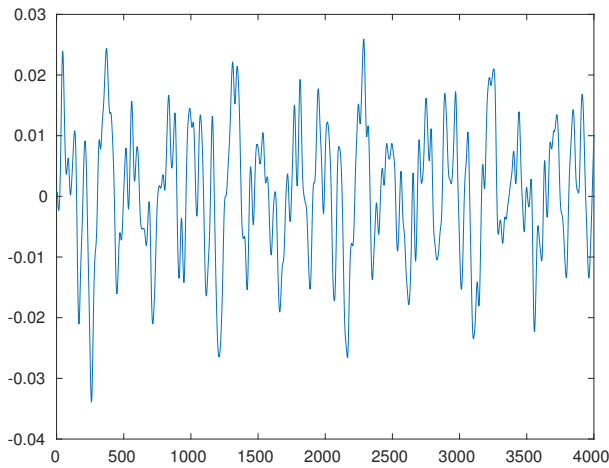


Figure 1.1: Wav input midi 50, 4024 points

II. SNAC FUNCTION AND MODIFIED CEPSTRUM

i. SNAC algorithm

The SNAC algorithm is based on autocorrelation, but the normalisation is done using the energy at each instant of calculation τ used for the autocorrelation. It's expressed by the following function:

$$n'(\tau) = \frac{2 \sum_{j=0}^{W-1-\tau} x_j x_{j+\tau}}{\sum_{j=0}^{W-1-\tau} (x_j^2 + x_{j+\tau}^2)}, \quad (1.1)$$

where W is the length of the windows, τ the discretised instant and x_j the sample to analyse.

The denominator as shown in the thesis [1] can be done in an iterative way:

$$\begin{aligned} m'(\tau) &= \sum_{j=0}^{W-1-\tau} (x_j^2 + x_{j+\tau}^2), \\ &= m'(\tau - 1) - x_{\tau-1}^2 - x_{W-\tau}^2, \end{aligned} \quad (1.2)$$

with $m'(0) = 2 \sum_{j=0}^{W-1} x_j^2$.

The figure 1.2 shows an example of result from a SNAC function. It's that a time domain signal centred around zero. It allows to define a peak between two zeros. The pitch of the signal is defined as the maximum peak, around $n=300$, thus the pitch is situated at $F_s/300$ which correspond to midi 50 as expected.

ii. Modified cepstrum

The modified cepstrum come also from the thesis of P.McLeod [1]. It results of the addition of 1 in the log part of the cepstrum avoiding that the small values affect the result of the cepstrum.

The modified cepstrum is expressed as:

$$mcep = FFT(\log(1 + abs(FFT(x)))) \quad (1.3)$$

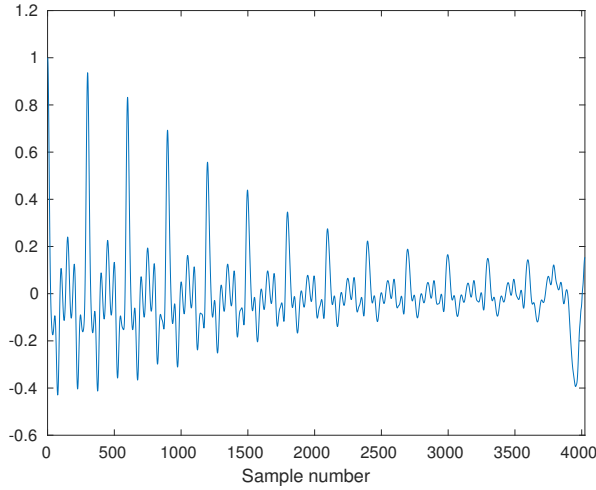


Figure 1.2: Example of result for SNAC algorithm, midi 50, 4024 pt

The figure 1.3 shows an example of result from a modified Cepstrum function. It's observed the same behaviour than the SNAC function, although there is a repetition of the half shape due to the FFT calculation.

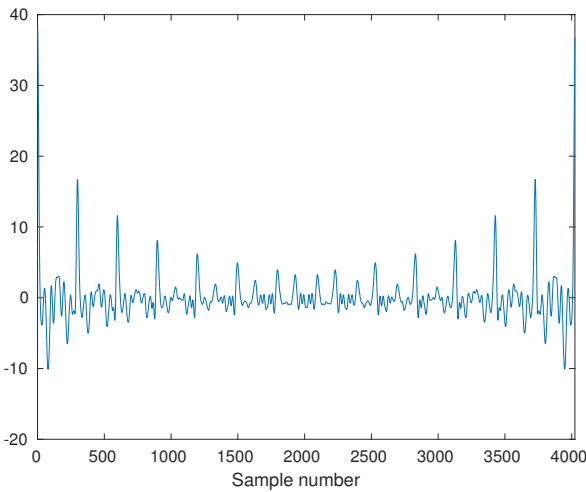


Figure 1.3: Example of result for modified Cepstrum algorithm, midi 50, 4024 pt

III. PRECONDITIONING THE INPUT SIGNAL

i. Window of signal

Instead of using a rectangular window it's used a Hann window which reduce the contribution from the edges, the hann windows is shown on figure 1.4.

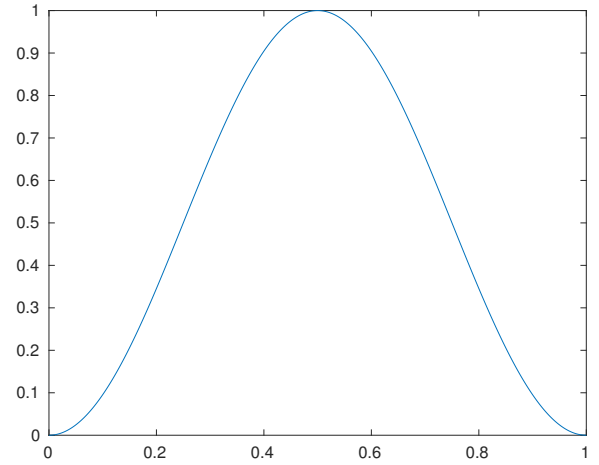


Figure 1.4: Hann windows used for preprocessing

ii. Outer medium ear filter

The ears have a special frequency shape on the audio range (20Hz-20kHz), thus the loudness of the sound depend on the frequency. The figure 1.5 shows the perceived equal loudness (phone, in red) as function of the frequency and the level (dB_{spl}).

It's proposed in P.McLeod to make a filter simulating this effect using a yulewalk filter and a second order low pass Butterworth filter, the resulting filter is shown on figure 1.6.

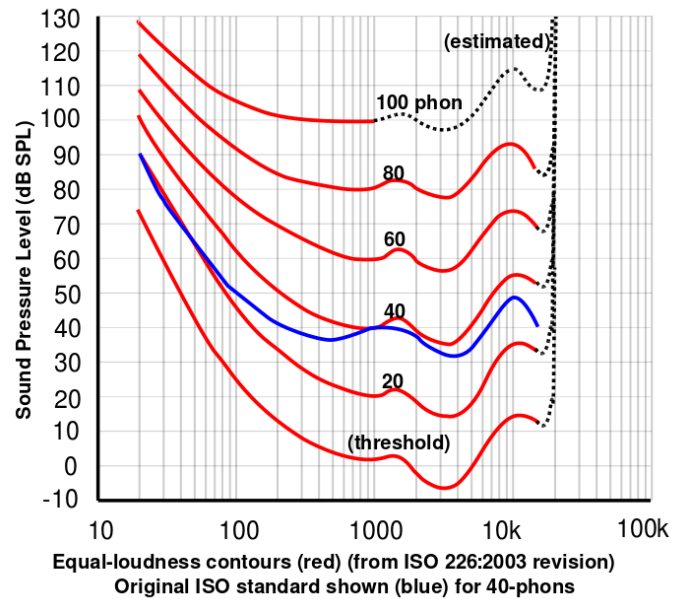


Figure 1.5: Perceived loudness of the sound

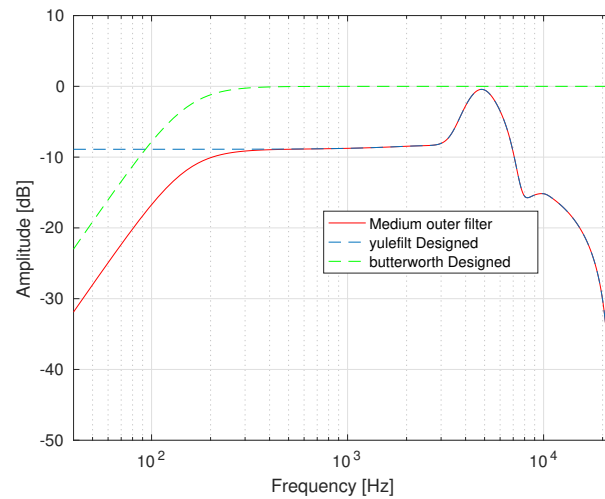


Figure 1.6: *Medium outer filter*

- iii. Aggregate Lag Domain (ALD)
- iv. Peak picking method

Do a graph recap
of the algorithm

IV. ASSERTION OF BOTH METHOD.

Chapter 2

Music note classification

Chapter 3

Music regeneration

I. CONCLUSION

Bibliography

- [1] McLeod Philip, *Fast, accurate pitch detection tools for music analysis*, Academisch proefschrift, 2009
, University of Otago. Department of Computer Science

[<http://www.piano-midi.de/mendelssohn.htm>]
