Signals and systems

Time and frequency analysis of music signals

Project report

Mattek4 G4-101a

Aalborg University
Department of Mathematical Sciences

In this project Python 2.7.11 has been applied for data processing and to draw graphs. The project is written in LaTeX.

AALBORG UNIVERSITY
STUDENT REPORT

Maths-Technology
Aalborg University
http://www.en.aau.dk/

Title:
Time and frequency analysis of music signals

Theme:
Signals and systems

Project Period:
Spring semester 2017

Project Group:
Mattek4 G4-101a

Participant(s):
Christian Hilligsøe Toft
Frederik Appel Vardinghus-Nielsen
Martin Kamp Dalgaard
Trine Nyholm Jensen

Supervisor(s):
Peter Koch
Henrik Garde

Copies: 4

Page Numbers: 14

Date of Completion:
Indsæt dato!

Synopsis:

indsæt

Contents

Todo list

Preface

Indsæt forord!

Aalborg University, February 14, 2017

Christian Hilligsøe Toft
cht15@student.aau.dk

Frederik Appel Vardinghus-Nielsen
fvardi15@student.aau.dk

Martin Kamp Dalgaard
mkda15@student.aau.dk

Trine Nyholm Jensen
trijen15@student.aau.dk

1 | Problem analysis

In this chapter an introduction will lead to a discussion of problems encountered by a musician when music is to be transcribed. The importance of these problems is analysed and the need for a solution assessed whereafter existing and possible solutions with roots in mathematics will be presented. The chapter will conclude with a *problem statement* which will form the basis for the rest of the project.

1.1 Introduction

Some musicians are capable of playing without reading music written on a sheet of music, and they invent new music through a creative process where they play by ear and do not have the need to write down their compositions. Creative thinking and processes are however interruptable and this can be problematic when the need for transcribing a composition - so as to remember or convey it to others - to a sheet of music arises. This may ruin the creative proces by exactly interrupting it because one needs to concentrate on the transcription.

To ease the creative process of a musician it is imaginable that some kind of automatic real time transcription of music to sheet music can be developed which will eliminate the problem of having to interrupt the aforementioned process.

1.2 Problem analysis

Although the first system of musical notation is the Sumerian system created 3,500 years ago music has existed for much longer. [7] This means that people have been playing music without any form of written music until then. Even with the creation of standardised musical notation musicianship on high level is possible without any form of education and/or skills in reading and writing music. As such there continue to exist musicians without aforementioned skills.

Inability to read or write music poses no problem in itself, as it doesn't necessarily hinder musical creativity. Problems however arise when the need to convey or remember musical compositions presents itself. Just as standardised languages make everyday communication and tasks easier a standardised musical notation is needed to convey compositions to others without anyone having to remember the compositions in their entirety. Standard musical notation has been created but has to be learned and research suggests that the ability to read and write music varies considerably from person to person and might even be affected by dyslexia. [4] In short people are genetically diverse when it comes to learning to read music.

If musical transcription in the middle of a creative process potentially interrupts said process the effect would be increasingly strengthened by the inability to quickly perform transcriptions which in the worst case scenario would require another person to do the transcription.

These consequences of the inability to read/write in musical notation may hinder the distribution of otherwise ingenious musical creations.

1.2.1 Existing and possible solutions

There has been undertaken plenty of research regarding automatic transcription of music to a note sheet and the material is to find in both books [5] and articles [2]. There also exists downloadable software which helps transcription by hand [6]. The extensive research done stems from the applications extending to other areas than just transcribing music for musicians - making computers participate with human performances is another imaginable application. Although the research is extensive "[···]the performance of transcription systems is still significantly below that of a human expert[···]" [1] and so there is room for improvement.

This project will focus on creating a solution with the use of mathematical tools to extract signal information from a digitalized analog signal of music so as to transcribe it to a note sheet. The viability of a solution in form of an algorithm based on the mathematical tools will be tested in a lab to illustrate the possibilities and limitations of said algorithm.

Her skal måske uddybes, når vi ved mere om vores løsning og metode.

1.2.2 Problem statement

The above analysis is summarized.
When musicians compose new music they firstly play but it stunts their creativity when they need to stop to transcribe the music. There is furthermore not necessarily any correlation between musical talent and the ability to transcribe music. This makes it advantageous for many musicians to use a program which automatically transcribes music to a sheet of music. This leads to the following problem statement:

> *How can an algorithm through the use of mathematical tools be designed to automatically transcribe music to a note sheet?*

Limitations

Given the time scope and the capability of collaborators of this project it is not expected to solve the above problem statement to full extent. Considering this a number of delimitations for the project will be decided upon in chapter 3.

2 | Music Theory

In this chapter basic music theory necessary for the understanding of the report is presented.

An example of a staff system with 15 notes is shown in figure 2.1. A note in such a system is generally a symbolic representation of a certain pitch, which is associated to a particular frequency (specified in hertz, Hz). Each note thus contains information about the frequency of the associated pitch and also about its duration in time. A note sheet is therefore actually a symbolic representation of a diagram over time and frequency, which is also called a spectrogram.

Figure 2.1: Example of notes in a staff system.

A staff system consists of 5 horizontal lines whereupon the notes are placed. Basically, only 12 key notes are repeated throughout the system but each with different frequencies. The key notes with the same names differ by an interval of n octaves for some integer n and are said to be octave equivalent. [8] The relation in frequency between two pitches A_0 and A_n are thus $A_n = 2^n A_0, n \in \mathbb{Z}$, and A_n is therefore n octaves above A_0.

The notes in the staff system are named by the letters A-G depending on their vertical position. After G the naming is repeated with the next A, which lies an octave above the previous one. Furthermore, between several of these pitches lie semitones, which are noted as e.g. $A^{\#}$. Table 2.1 shows all pitches in the interval A_3-A_4.

To be able to distinguish between octave equivalent pitches they are e.g. noted as A_4 and A_5. A_4 is also called the concert pitch and is normally chosen to be 440 Hz by definition but may vary by ±3 Hz depending on who's playing. All other pitches can be defined from the chosen frequency of A_4 since it is the 12^{th} key note after A_3, whose frequency is half the size of A_4's. The relation in frequency between two neighbouring pitches, e.g. A and $A^{\#}$, is therefore $A^{\#} = \sqrt[12]{2} \cdot A$. [8] The frequencies of the pitches in the interval A_3-A_4 rounded off to the nearest integer is shown in table 2.1.

When a pitch is played the pitch's overtones will sound along with the actual pitch. The frequency of a pitch's overtone is an integer multiple of the original pitch's frequency, and this phenomena is therefore among others reflected through the octave equivalent pitches of the original one. However, the pitch A_2 with frequency 110 Hz also has the overtone E_3 (330 Hz) along with the octave equivalent pitches A_3 (220 Hz), A_4 (440 Hz) and A_5 (880 Hz). If several instruments play at the same time all the original pitches and overtones are mixed, which is one of several reasons why it is difficult to translate music played by several instruments into a single note sheet. On the other hand, if only one instrument is playing, the overtones are not mixed together, and the pitch is thus easily distinguished from the overtones. The pitch in question will then usually be the one with the lowest frequency. The overtones therefore complicate the determination of the original pitch, but the determination is not impossible if the pitch e.g. is played alone in an anechoic room. This is shown in figure (insert reference to figure below).

Table 2.1: Frequencies of the pitches A_3-A_4 shown in Hz and rounded off to nearest integer.

Pitch	A_3	$A_3{}^{\#}$	B_3	C_3	$C_3{}^{\#}$	D_3	$D_3{}^{\#}$	E_3	F_3	$F_3{}^{\#}$	G_3	$G_3{}^{\#}$	A_4
Freq	220	233	247	262	277	294	311	330	349	370	392	415	440

Insert spectogram showing a single pitch with overtones here.

Playing a pitch alone in an anechoic room is a way of minimizing noise factors. There are many different noise sources from e.g. the surrounding environment, electrical equipment, other people and even the instrument playing itself since it is difficult to make a pure tone by an acoustic sound source such as a guitar. A pure tone is a sinusoidal waveform consisting of a single frequency and may therefore be difficult to play on an instrument. [3] Usually, sound is reflected off the walls in a room which is also a source of noise. This is a form of folding and is minimized in the anechoic room because sound is absorbed by the walls. Moreover, due to the construction of the anechoic room, noise from e.g. bypassing cars is also minimized, and the sound may be recorded with a minimum of hardware which otherwise may also produce noise.

The music used in this project will therefore be recorded in an anechoic room because the noise is minimized. It is not expected that musicians using the system in the future have access to an anechoic room as well but if the system doesn't work with sounds recorded in the anechoic room then it with most certainty doesn't work at other places neither. However, in order to reproduce the conditions of a typical musician working around other people, background noise can also be made in the anechoic room but should of course not drown the music. This form of noise is additive whereas e.g. noise reflected off walls as mentioned above is multiplicative. In general, multiplicative noise depends on the state of the system whereas additive noise does not. Therefore, the output $x[n]$ of the sampled data $s[n]$ corrupted by the additive noise $a[n]$ is $x[n] = s[n] + a[n]$ whereas the equation for the multiplicative noise $m[n]$ is $x[n] = s[n]m[n]$.

Comment on this chapter: does it need a conclusion? Or is there something else missing? Which one of the words "pitch" and "tone" is better to use? According to Wikipedia, "tone" is the best word (`https://en.wikipedia.org/wiki/Musical_tone`) but I personally prefer "pitch" over "tone". ®

3 | System model and requierments

In this chapter a model of and requirements to the system is specified on a superficial level. These specifications forms a basis of the further work.

Due to the presented problem statement and delimitation of the project the purpose of the application is basically to transform a sound into a spectrogram showing the frequency due to the sound. This is illustrated by figure 3.1. In order for an application to perform this

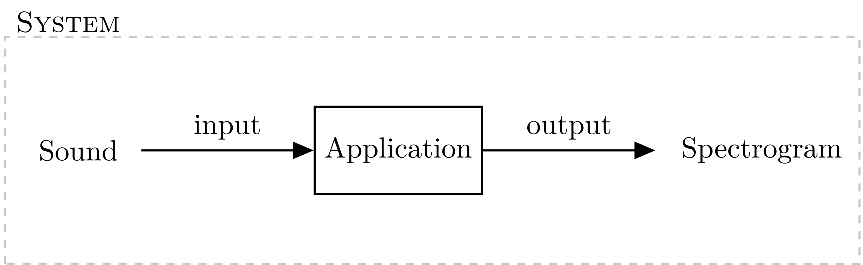

Figure 3.1: Basic block-diagram of the system

transformation an analogue sound has to be converted into a discrete time signal, in order for a computer to process the signal into the wanted outcome.
By this the system is fundamentally based upon theory about discrete time systems, including analysis and processing of such systems and the disposal data.
In the following sections specifications and requirements are specified for the three parts of the system.

3.1 Input

To process an analogue sound in a computer the sound has to be digitalized, this is done by an analogue to digital converter(ADC). An ADC basically consistent of sampling of the signal done by a sample and hold unit(S/H) followed by quantifying of the samples. Further the signal has to be stored as a datafile suited to the application. This is illustrated by figure 3.2

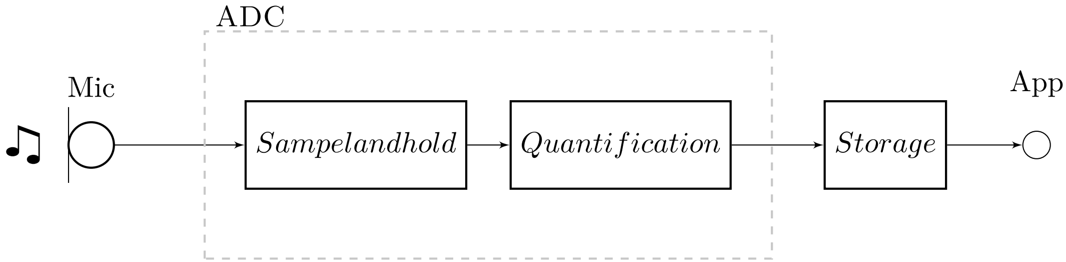

Figure 3.2: Basic block-diagram illustrating an ADC where the digital output will be the input to the application

3.2 Output

3.3 Application

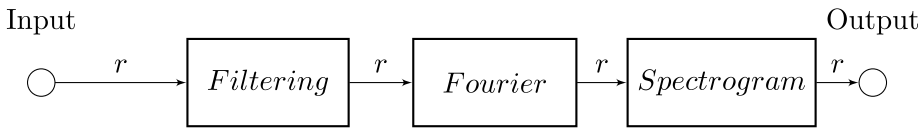

Figure 3.3: fdklga

4 | Overskrift

About noise from chapter 2: ®

Playing a pitch alone in an anechoic room is a way of minimizing noise factors. There are many different noise sources from e.g. the surrounding environment, electrical equipment, other people and even the instrument playing itself since it is difficult to make a pure tone by an acoustic sound source such as a guitar. A pure tone is a sinusoidal waveform consisting of a single frequency and may therefore be difficult to play on an instrument. [3] Usually, sound is reflected off the walls in a room which is also a source of noise. This is a form of folding and is minimized in the anechoic room because sound is absorbed by the walls. Moreover, due to the construction of the anechoic room, noise from e.g. bypassing cars is also minimized, and the sound may be recorded with a minimum of hardware which otherwise may also produce noise.

The music used in this project will therefore be recorded in an anechoic room because the noise is minimized. It is not expected that musicians using the system in the future have access to an anechoic room as well but if the system doesn't work with sounds recorded in the anechoic room then it with most certainty doesn't work at other places neither. However, in order to reproduce the conditions of a typical musician working around other people, background noise can also be made in the anechoic room but should of course not drown the music. This form of noise is additive whereas e.g. noise reflected off walls as mentioned above is multiplicative. In general, multiplicative noise depends on the state of the system whereas additive noise does not. Therefore, the output $x[n]$ of the sampled data $s[n]$ corrupted by the additive noise $a[n]$ is $x[n] = s[n] + a[n]$ whereas the equation for the multiplicative noise $m[n]$ is $x[n] = s[n]m[n]$.

5 | Overskrift

6 | Overskrift

Bibliography

[1] Benetos, Emmanouil et al. "Automatic Music Transcription: Challenges and Future Directions". In: (2013). URL: http://openaccess.city.ac.uk/2524/.

[2] Camgil, Ali Taylan, Kappen, Bert, and Barber, David. "A Generative Model for Music Transcription". In: (). URL: https://infoscience.epfl.ch/record/83228/files/barber-idiap-rr-05-89.pdf.

[3] Crocker, Malcolm J. and Ffowcs-Williams, John E. *Acoustic Noise.* AccessScience, McGraw-Hill Education. Visited 09-02-2017. 2014. URL: http://www.accessscience.com/content/acoustic-noise/006100.

[4] Ganschow, Leonore, Lloyd-Jones, Jenafer, and Miles, T.R. *Dyslexia and Musical Notation.* Springer, 1994. URL: http://www.jstor.org/stable/23769692.

[5] Klapuri, Anssi and Davy, Manuel. *Signal Processing Methods for Music Transcription.* Springer, 2006.

[6] Software, Seventh String. *Transcribe!* URL: https://www.seventhstring.com/xscribe/overview.html.

[7] Wallin, Nils L., erker, Björn, and Brown, Steven. *The Origins of Music.* The MIT Press, 2000.

[8] Wright, David. *Mathematics and Music.* Visited 08-02-2017. 2009. URL: http://www.math.wustl.edu/~wright/Math109/00Book.pdf.

A | Overskrift

B | Overskrift