Спектрален анализ на звуков сигнал

Валентин Латунов

Abstract

Човешкият мозък възприема звука като електрически импулси от невроните в слуховия апарат, но го интерпретира въз основа на по-сложни метрики. Тон, тембър, ритъм и хармоничност са някои от свойствата на звука. В този доклад ще разгледам метод за тяхното извличане чрез прилагане на дискретна трансформация на Фурие върху сигнал.

1. Introduction

Въпреки, че обученото ухо може до определена степен да улавя тези характеристики на звука, компютъра вижда само поредица от числа (масив от данни). Проблемът в тях е, че те не са добра мярка за повечето полезни или отличителни свойсвта на звука. Задачата е от тях, чрез различно представяне на информацията, да извлечем отделните качества. Решението, което ще разгледам, е чрез разлагането на звуковия сигнал в домейн на честоти и амплитуди и генерирането на характеризиращо изображение. Това дава възможност да се прилагат разпознавателни техники за извличане на информация (броене, статистики, machine learning и др).

2. Methods and Materials

2.1. *Формат на данните*

Сигналът е представен като масив от реални числа в интервала $[-1, 1]$. Те са семплирани с честота 48000 Hz.

2.2. *Математически формули*

Разлагането на Фурие за функция f(x) има вида:

$$f(x) = \tfrac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nx) + b_n \sin(nx), \qquad a_k, b_k \in R$$

Съкратено може да се запише:

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}, \qquad c_k \in C$$

Коефициентите $\mathrm{c_k}$ се намират чрез формулата:

$c_k = \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-ikx}dx$

Тъй като сигналът е дискретна функция, дефинирана в N точки, то:

$$\int_0^{2\pi} f(x)e^{-ikx}dx = \sum_{j=0}^{N-1}\int_{x_j}^{x_{j+1}} f(x)e^{-ikx}dx \approx \sum_{j=0}^{N-1}(x_{j+1}-x_j)\frac{f(x_{j+1})e^{-ikx_{j+1}} + f(x_j)e^{-ikx_j}}{2}$$

Освен това, точките x_j са равно отдалечени.

$x_j = \frac{2\pi j}{N} \Longrightarrow x_{j+1} - x_j = \frac{2\pi(j+1-j)}{N} = \frac{2\pi}{N}$

Тогава:

$$\sum_{j=0}^{N-1}(x_{j+1}-x_j)\frac{f(x_{j+1})e^{-ikx_{j+1}} + f(x_j)e^{-ikx_j}}{2} = \sum_{j=0}^{N-1}(\frac{2\pi}{N})\frac{f(x_{j+1})e^{-ikx_{j+1}} + f(x_j)e^{-ikx_j}}{2} =$$

$$\frac{\pi}{N}(f(x_0) + 2\sum_{j=1}^{N-1} f(x_j)e^{\frac{-2\pi ikj}{N}} + f(x_N))$$

Тъй като смятаме, че сигналът е периодична функция, $f(x_0) = f(x_N)$.

$$\frac{\pi}{N}(f(x_0) + 2\sum_{j=1}^{N-1} f(x_j)e^{\frac{-2\pi ikj}{N}} + f(x_N)) = \frac{2\pi}{N}\sum_{j=0}^{N-1} f(x_j)e^{\frac{-2\pi ikj}{N}}$$

Окончателно за коефициентите c_k получаваме:

$$c_k = \frac{1}{N}\sum_{j=0}^{N-1} f(x_j)e^{\frac{-2\pi ikj}{N}}$$

За пресмятането на N коефициента за необходими N^2 операции. Това може да се подобри като се възползваме от факта, че:

$$c_k = \frac{1}{2}[\frac{1}{M}\sum_{j=0}^{M-1} f(x_{2j})e^{\frac{-2\pi ikj}{M}} + (\frac{1}{M}\sum_{j=0}^{M-1} f(x_{2j+1})e^{\frac{-2\pi ikj}{M}})e^{\frac{-2\pi ik}{N}}], \qquad M = \frac{N}{2}$$

Това ни казва, че разлагането на Фурие от N-ти ред може да се изрази като две разлагания от $\frac{N}{2}$-ти ред върху четните и нечетните индекси на стойностите на функцията. Тогава, ако $N = 2^p$, броя операции за да пресметнем всички коефициенти е:

$T(N) = 2T(\frac{N}{2}) + \Theta(N) \asymp N\log(N)$

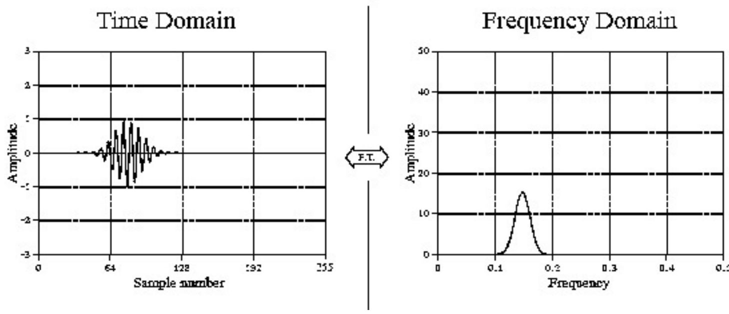
2.3. Използване на трансформацията на Фурие за генериране на спектрограма

След прилагане на трансформацията върху сигнала получаваме нов масив от вече комплексни числа, реалните и имагинерните части на които съответсват на косинусовата и синусовата компонента от дадена честота. Тъй като искаме да изразим амплитудата за дадеда честота използваме равенството:

$$x\cos(\theta) + y\sin(\theta) = \sqrt{x^2 + y^2}\cos(\theta - \arctan(\frac{y}{x}))$$

Така лесно се вижда, че амплитудата A $= \sqrt{x^2 + y^2}$.

Недостигът на метода е, че се губи информация за времето.



Фигура 1: Смяна на домейн

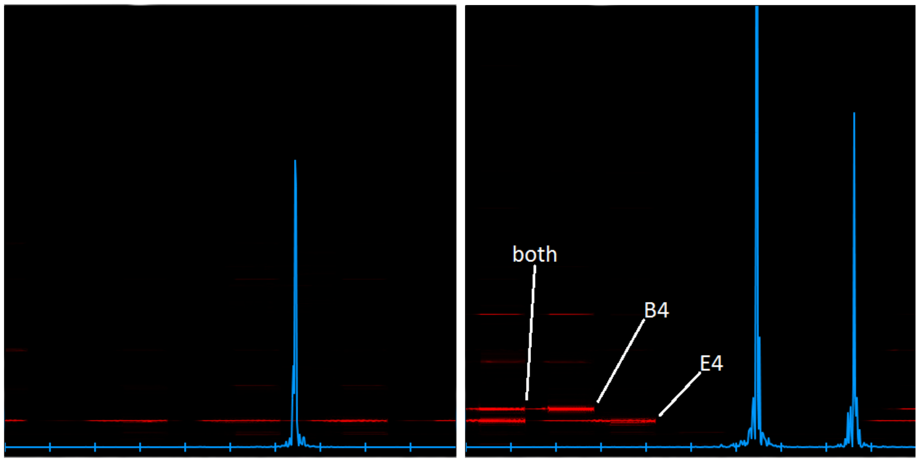
В честотния домейн виждаме кои честоти съставят сигнала, но не може да се определи кога става промяна в амплитудите, ако има такава. Освен това не е практично да се прави трансформация върху големи по дължина сигнали. Затова сигналът се разбива на по-малки участъци, които поотделно трансформираме. Така до определена степен можем да кажем кога се среща дадена честота.

Пресмятаме квадратите на амплитудите за всяко парче. Това ни дава двумерна таблица от стойности - спектрограмата. За да няма големи разлики между съседни клетки от таблицата е честа практика да се разглеждата застъпващи се парчета.

3. Results and discussions

Метода позволява свобода в някои от параметрите си. Зависимо от тях се получават различни резултати.

Един от най-важните е дължината на отделните парчета. От нея зависи максималната честота, която може се различи. За N-ти ред трансформация можем да различим $\frac{N}{2}$ честоти (от 0 до $\frac{N}{2}$-1). Освен това, колкото по-дълго е парчето, толкова по-малко информация имаме за това кога някоя честота се проявява. Тази зависимост е свързана със Съотношението на неопределеност на Хайзенберг. Накратко казано - не можем да получим произволно добра резолюция едновременно по отношение на време и честота. Обикновено се избира подходящ компромис за конкретния проблем.



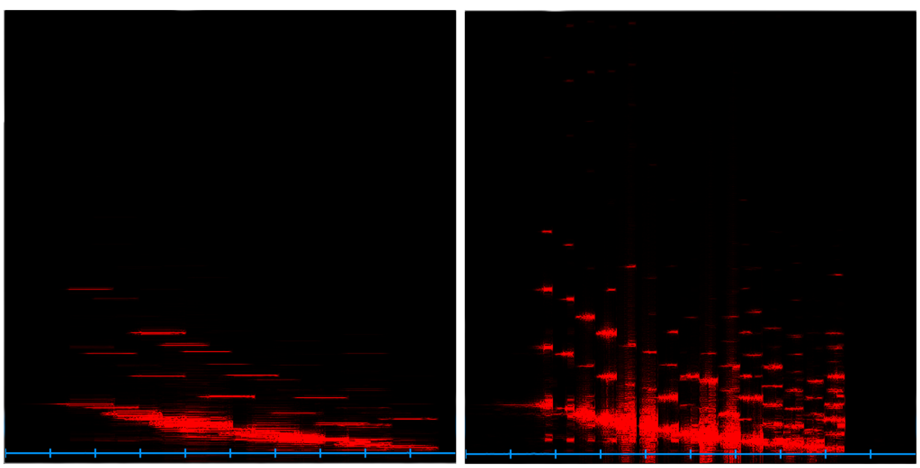
Фигура 2: Спектрограма на две ноти E4, B4

За спектрограмата във фиг. 2 са използвани следните параметри:

Брой семпъли = 8192

Дължина на прозореца = 1 сек.

В червено е изрисувана спектрограмата. Синята крива е най-левия стълб от нея. Вертикалните черти разграгяват честоти през 50. Яркостта на цвета съответства на квадрата на амплитудата.



Фигура 3: Спектрограма на арпежиран до-мажор

За спектрограмата във фиг. 3 са използвани следните параметри:

Брой семпъли = 8192

Дължина на прозореца = 1 сек. (ляво) / 0,17 сек. (дясно)

Тук се вижда разликата в резолюцията по отношение на време и честота. По-дългият прозорец позволява за по-точно разграничаване на честотите, а по-късият - на времето.

4. Conclusions

При правилно настройване този инструмент може да ни даде добра визуална представа за структурата на един сигнал - дали това е шум, музика, реч или друго. Нещо повече - той дава основата за извличане на качествата на сигнала. За всеки случай има конкретни техники и трикове за прилагане, които подобряват желания резултат.

Този метод не е единствения разработен. Вместо фиксирана резолюция по отношение на времето и честотата, може да се постигне променлива с помощта на уейвлетни разлагания.