

СИСТЕМА ЗА ИЗВЛИЧАНЕ НА ИНФОРМАЦИЯ ОТ ЗВУК В РЕАЛНО ВРЕМЕ

Виктор Матански

Резюме: Извличането на информация от звук привлича огромно внимание през последните 20 години. Изследванията в тази област непрекъснато нарастват и намират все повече и повече място в ежедневието на хората чрез различни технологии. От най-очевидното приложение на добиване на информация от музика, до по-специфичното извличане на данни отнасящи се до вибрационната характеристика на движещи се механични части, човешки органи, сгради и конструкции, както и инфраструктура. Въпреки това, откриването на шаблони в информацията е умение развивано от зората на човешката еволюция до ден днешен. В тази статия ще представя идея за система за извличане на информация в реално време, в която основния акцент е добиването на данни от звук. Целта на системата е да предостави ефикасен начин за използване на алгоритми за извличане на данни и тяхната комбинация, както и начин за съхраняване на получената информация във формата на обекти, които са удобни за допълнителна обработка и използване.

Ключови думи: анализ на звук, класификация, извличане на данни, машинно обучение

1. Въведение

Областта на извличане на информация се отнася до много повече от изолираното извличане на данни от текст. През последните десетилетия напредъка в тази сфера е дал множество полезни технологии и системи за работа с информация, като по този начин е оказал огромно влияние върху света. Един от ярките примери за такива системи са Уеб базираните търсачки като Google и Yahoo, които са основен фактор в интернет пространството. След успехът в работата с текст тези системи вече започват да боравят и с други видове информация като изображения, видео и звук (Meurer).

В днешно време начинът, по който се борави със записан звук и в частност с музика, начините по които се създават, съхраняват и възприемат дадени музикални произведения е почти изцяло дигитален. Компютърните технологии изместват реалните звукови сигнали като жива, динамична музика и дигитализират самото действие на слушане. Можем да направим аналогия с времето преди ХХ век, когато понятието за звук и музика е можело да бъде възприето единствено при тяхното директно възпроизвеждане, което е строго ограничено от местоположение и време. Днес, голяма част от звуковата информация, била тя звуци на живи същества, природата или музикални произведения, налична до момента може да бъде намерена в дигитален формат (Smith). Тази информация може да бъде записана и съхранявана по различни множество начини и следователно може да бъде възпроизведена и мултиплицирана без ограничение.

Извличането на информация от музика представлява дисциплина, която е привлякла вниманието на множество изследователи от различни сфери като компютърни технологии, информационни науки, психология, музика и акустика. Извличането, анализът и организацията на информация представляват основите три точки в тази дисциплина. Основната цел в тази сфера е развитието на ефективни методи за употреба и работа със звук и разширяването на разбирането и полезността на музикалните данни, чрез изследване, развитие и приложение на изчислителни модели и средства (Wiering).

2. Проблематика

Въпреки напредъка на дисциплината, като задача алгоритмичният анализ на звук остава все още не разрешен. Това може да се дължи на причината, че звукът и особено музиката като концепция представляват идея, която е отвъд информационното разбиране на човек. Нека разгледаме следния пример. При извличането на информация, самата дума „информация“ може да се определи като структура от данни, която описва материалния свят (обекти и явления) с цел облагородяване на знанието за света. Когато разглеждаме новини за света, учебни пособия, навигационни карти или упътвания информацията е обективна и строго определена. Но когато разглеждаме литературни произведения информацията придобива друг, абстрактен смисъл и освен семантичните данни се появява друг аспект, който е чисто субективен и предизвиква проблематика при опита за строга дефиниция. Затова може да се каже, че при опита за дефиниция на подобен вид информация няма как да се изгуби част от нейния аспект, тъй като тя не е строго определена. По същия начин може да се разгледа музиката като изкуство или звукът като усещане. Извличането на данни, употребата им като обекти и създаването на връзки между тях може да спомогне разрешаването на този проблем, като позволи различни комбинации на данните и лесната работа с тях.

3. Видове звукова информация

Механични или символични данни

Един алгоритъм за анализ на звукова информация може да обработва поток от данни, било то като запис или в реално време, извличайки съдържателна информация като това къде е разположен източника на звук, дали той притежава дублиращи се свойства, какви са честотните характеристики и т.н. Този вид данни са обективни, тъй като са материални, строго дефинирани в своето описание и разглеждат механични качества. В същото време от звук и музика могат да се извлекат други не механични данни като жанр и стил, изпълнител, настроение, които в своята дефиниция са абстрактни и се състоят от символична информация.

Дескриптори високо ниво

Интуитивната начална точна за извличане на информация по съдържателен път представлява използването на мелодия или хармония за описание на музика. В зората на тази сфера, са били разработени множество системи, които са били базирани на заявка чрез наподобяване на мелодия (humming) (Ghias, 1995). Идеята е била да се търси в база данни съдържаща полифонични записи. Потребителят е можел да намери това, което търси опитвайки се да възпроизведе мелодията. Доста подобни системи вече са на комерсиалния пазар (SoundHound, Shazam) и имат голям успех (Heinrich). Дескрипторите от високо ниво представляват знание, което е от няколко типа едновременно. Това знание може да е сложно и част от него дори непонятна за потребителя.

Въпреки, че е интуитивен процес извличането на мелодия от полифонични записи, т.е. множество инструменти възпроизвеждащи различни мелодии едновременно, тепърва предстои пълната имплементация на подобен алгоритъм. Изненадващо е, че не само е трудно да се извлече мелодията от аудио сигнала, но и дори и символична репрезентация като MIDI файлове. Същото важи и за други идеи от високо ниво като ритъм, тембър и хармония. Следователно, извличането на музикално съдържание от високо ниво е част от тази дисциплина и подлежи на усилени

изследвания. Крайната цел е тази информация да се постави в стандарт и да може да се използва за търсене и класификация.

Дескриптори от ниско ниво

Автоматичното извличане на данни от звук може да бъде разгледано и по друг начин. Има предположения, че извличането на дескриптори от ниско ниво и използването им като средства за добиване на информация е за предпочитане пред директното извличане на дескриптори от високо ниво. Освен това, възможността за извличане на дескриптори от ниско към високо ниво също е обект на изследване. Ограниченията в един вид алгоритми могат да бъдат преодолявани използвайки различни такива или няколко в комбинация.

Дескрипторите от ниско ниво представляват качества на звука, които съдържат измерима информация за музикално произведение. Те също така съдържат неприсъщи данни, поради трудността да се измери само един аспект на звуковия запис, така че е налична обмяна между дескрипторите от различни нива.

Дескрипторите от ниско ниво са разделени по три различни начина:

- сегментации базирани на кадри (периодично семплиране на интервали от 10 милисекунди до 1000);
- сегментации базирани на ритъм (свойствата на записа са разделени на ритмични граници);
- статистически изчисления, които съставят вероятното разпределение на свойствата (модел) (Valchanov et. al., 2010), (Пиев, 2006).

4. Основи на системата

Основната цел на системата е да се създаде средство, с което да се извлича максимално ефикасно информация от звук в реално време. Поради вида на информацията, която представлява комбинация от механични и символични данни, тази цел е трудна за постигане. За да се елиминира максимално тази трудност трябва да се намери решение на субективния фактор съдържащ се в символичните данни. Това може да се случи като се въведе максимално гъвкава функционалност, която да може да бъде калибрирана спрямо спецификата на звука, който бива обработван в даден момент. Внедряването на техники от машинното обучение и изкуствения интелект са необходим фактор за последващото развитие на системата.

Системата може да бъде разгледана като три отделни модула:

- Модул за входящи данни – отговарящ за сигнала, който в следствие ще бъде обработван. Звукът, като физическо явление представлява аналогов елемент, но за да бъде обработен от компютър, то той трябва да бъде преобразуван в дигитален формат.
- Модул за обработка и анализ на данни – този модул работи с вече преобразуван, дигитален звуков сигнал. Вече разполагайки с дигитални данни, той може да използва един или повече алгоритми за извличане на информация. Възможните комбинации могат да бъдат определени автоматично или могат да бъдат избрани и допълнително калибрирани от оператор. При наличие на невалидни данни от една комбинация от алгоритми може да бъде приложена друга, алтернативна за случая такава.
- Модул за съхранение на данни – след като вече звуковата информация е била първо преобразувана от аналогов към дигитален сигнал в първия модул и след

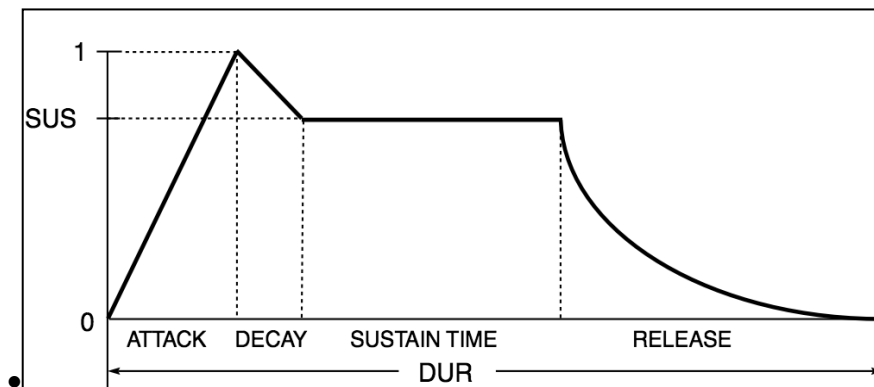
това от дигитален сигнал в качествени данни, то тя следва да бъде съхранена за по-нататъшна употреба. Важно е да се отбележи нуждата от съхраняване на информацията от всички нива на системата. Това е така, защото за да се създаде максимално ефикасна система за извличане на данни от звук е нужна подробна база за сравнение, разполагаща с всички стадии на информацията, която се обработва.

Нека проследим точно как протича процесът на пренос на информация от звук, през дигитализация до съхранение.

На първо място, какво представлява звукът? Звукът представлява механична вибрация, която се движи в плътна среда (не вакуум). Когато звукът е само един (монофоничен), той представлява една единствена вибрация, но когато звуците са повече от един (полифоничен), то броят на вибрациите се увеличава.

Как протича съществуването на една такава вибрация? Звукът притежава няколко основни компонента. Тези компоненти в акустиката представляват преходни процеси и се наричат attack, decay, sustain, release (атака, разпад, задържане, отпускане), които се разполагат в определен период от време (ADSR) (De Leon).

- Атака – времето, за което нивото на звука възлиза от 0 до своя връх, т.е. начало на възпроизвеждане на вибрация.
- Разпад – времето, за което нивото на звука преминава от своя връх до определеното време за задържане.
- Задържане – времето, за което основната звукова честота функционира.
- Отпускане – времето, за което след приключване на възпроизвеждането на звука нивото на вибрация достига до 0.



Фигура 1. Визуализация на различните компоненти на ADSR

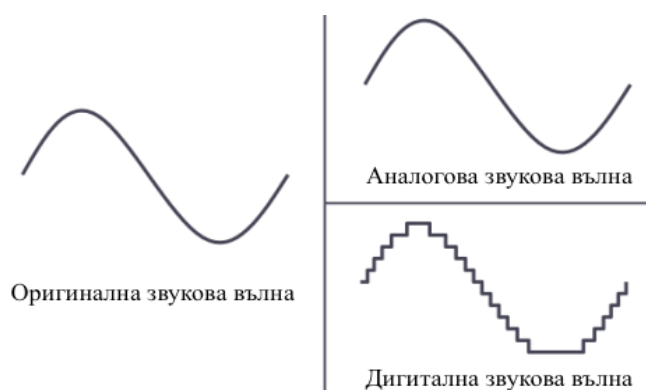
Всеки един звук или комбинация от звуци се развива във времето, а времето като фактор разполага с посока. Преходните процеси (ADSR) се случват във времето и могат да бъдат разгледани като негова функция. Следователно тези няколко елемента описват живота на една звукова вибрация. След като разполагаме с тази концепция, нека разгледаме по какъв начин можем да обработваме звуковата информация.

От аналогов към дигитален сигнал

Звукът, като механична вълна, предизвиква промяна във въздушното налягане. Човешката слухова система възприема тази промяна основно чрез граничната част между външното и средното ухо (тъпанчето). От там нерви в слуховия апарат реагират на вибрациите и изпращат електрически импулси към мозъка, който в последствие ги обработва и реагира на тях. Всеки един звук, бил той самостоятелен или комплексен, представлява промяна във въздушното налягане.

Подобно на биологичния слухов апарат, в аналоговите системи тези промени се улавят чрез записващи устройства (микрофони) и се превръщат в електрическо напрежение. Тъй като това представлява продължителен процес аналоговите сигнали се наричат продължителни или времеви сигнали, а амплитуда се нарича нивото на промяна в определен момент (положителна или отрицателна). И обратно: когато един сигнал се възпроизвежда чрез говорители, мембраната на тези говорители се движи подобно на вибрацията, която предизвиква промяна във въздушното налягане.

Звуковите записи имат два основни типа: аналогов и дигитален. Аналоговите записи се отнасят до звук записан чрез използването на методи, които копират оригиналната звукова вълна. Аудио касети и грамофонни плочи са пример за подобен вид аналогови носители. Дигиталните записи представляват извадки от оригиналната звукова вълна с определена характеристика. Компакт дискове и wav/mp3 са пример за дигитални носители.



Фигура 2. Сравнение на различните видове звуков сигнал

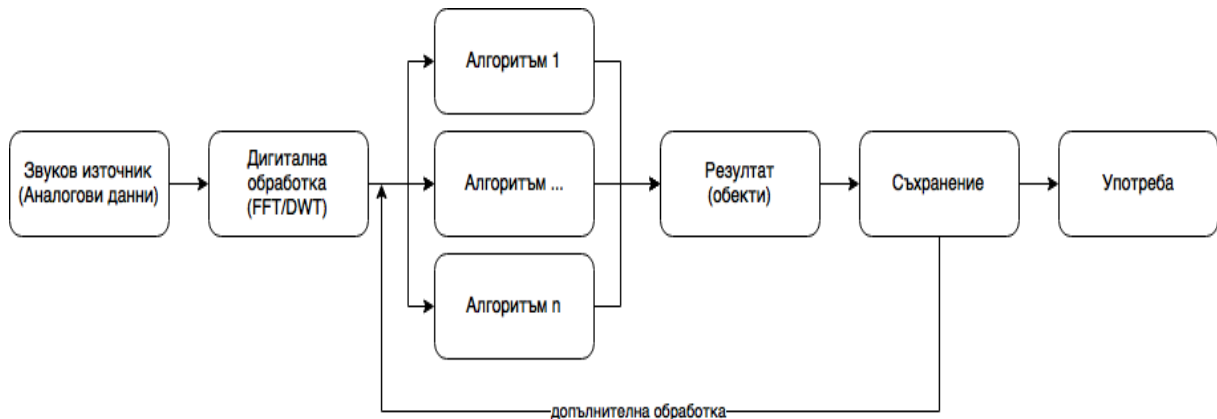
На фигура 2 се вижда как аналоговата звукова вълна е идентично копие на оригиналната звукова вълна, докато дигиталната звукова вълна представлява множество извадки от оригиналната звукова вълна. Качеството на аналоговите записи зависи от чувствителността на екипировката и средата, в която е направен записа. Качеството на дигиталните записи зависи от честотата на тези извадки, но въпреки това дори и най-новите дигитални технологии не успяват да пресъздадат качеството на звукови вълни.

Поради невъзможността на компютърните технологии да боравят директно с продължителни сигнали, този вид информация трябва предварително да бъде превърнат в поредица от стойности. Процесът на преобразуване на продължителен сигнал в поредица от стойности се нарича дискретизация (sampling). След като продължителното време се разпредели на кратки интервали, то вече те могат да бъдат представени за директна обработка от компютър. Броят на интервалите във времето се наричат честота на дискретизация, която се измерва в херци. Тъй като накъсването на продължителен сигнал на части предизвиква възможност за грешка, то при по-висока честота на дискретизация тази възможност намалее. След дискретизация, аналоговият сигнал се превръща в поредица от амплитудни стойности, които могат да бъдат съхранени и обработени от компютър.

След като разгледахме начина, по който съществува една звукова вибрация можем да разпределим функциите на системата спрямо различните модули.

- Модул за входящи данни – обработка на аналогов сигнал и преобразуването му в дигитален (Звуков източник, Дигитална обработка).
- Модул за обработка и анализ на данни – работа с дигитален сигнал чрез различни функции и извличане на звуци като обекти (Алгоритми, Обекти).

- Модул за съхранение на данни – запис на обектите създадени в модула за обработка и анализ на данни (Съхранение).



Фигура 3. Архитектура на системата за извличане на информация от звук

5. Прототип

За създаването на прототип на системата е използван програмният език Processing, който е базиран на Java и разполага със своя собствена среда за разработка (IDE). Езикът е избран пред други алтернативи поради възможността за бърза разработка на програми с визуален контекст (Processing IDE). Обработка на звук се извършва чрез библиотеката Minim, която използва JavaSound API. Библиотеката разполага с базови функции за работа със звук, които служат за основа на прототипа (Minim Libr). Допълнителен контрол в реално време се извършва чрез библиотеката MidiBus, която чрез MIDI протокол разрешава лесен и удобен контрол над параметри от системата, в реално време (MidiBus Library).

Част 1: Реализация на модул за входящи данни

За получаването на данни в реално време се използва външен микрофон. Информацията от него се обработва като се създава обект с име audioSource, в режим Stereo и честота на дискретизация от 1024. След като вече сигналът е дискретизиран, към него се прилага бързо преобразуване на Фурие (FFT) и той се превръща от времеви сигнал в честотен такъв. След това честотния диапазон от 0 до 22.000 херца се разделя на 50 равни части, с цел по-ефикасна обработка и пестене на ресурси.

```

Minim      minim;
AudioInput microphoneInput;
FFT        fft;
setupModule1()
{
  minim = new Minim(this);
  // създаване на вход за аудио данни чрез използването на външен микрофон
  audioSource = minim.getLineIn(stereo, 1024);
  // създаване на дигитална репрезентация на звуковия сигнал във времеви
буфер
  fft = new FFT( audioSource.bufferSize(), audioSource.sampleRate() );
  // разделяне на честотите от 0 до 22.000 херца на 50 равни части
  fft.linAverages( 50 );
}
  
```

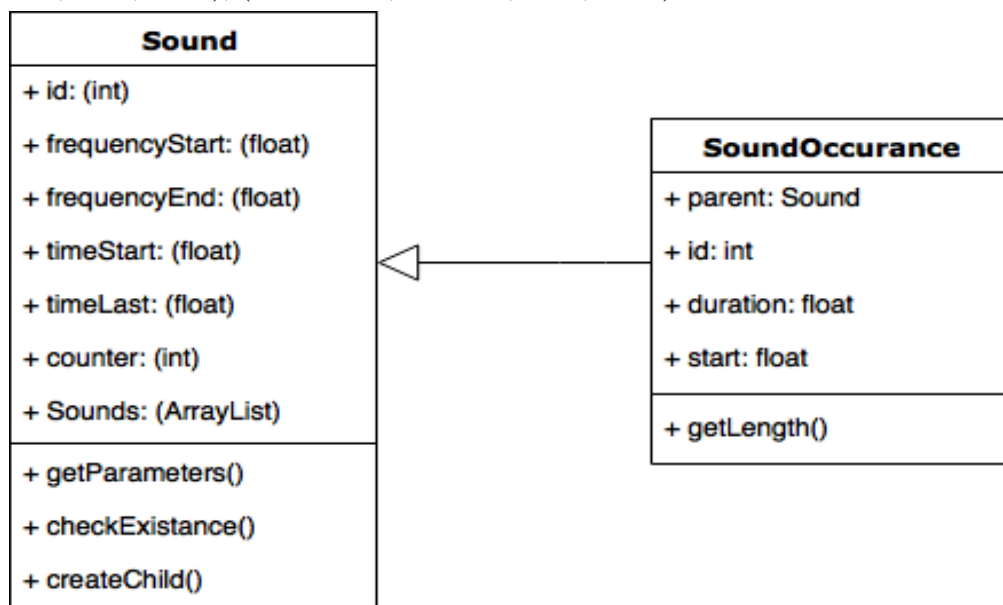
Част 2: Имплементиране на алгоритъм за извличане на данни

След като вече сигналът е дискретизиран, върху него е приложена функция на Фурие (FFT) и той бива превърнат в сигнал от честотната равнина. Като такъв, сигналът придобива амплитудна стойност за всяка една честота в диапазона от 0 херца до 22.000 херца (Bello et. al.).

В по-горните части бе описано, какво представлява един звук. За да се засече наличие на звук е създадена функцията `detectNotes(FFT fftObj)`, която работи по следния начин:

- първо се изчислява средна амплитудна стойност (`float averageFrequencyAmplitude`) за честотите от 0 до 4000 херца и към тази стойност се добавя параметър `float averageDifference`, който се контролира от оператор, за да се калибрира чувствителността на измерване;
- след това се проверяват честотите в същия диапазон за наличност на амплитудна стойност по-голяма от `averageFrequencyAmplitude`. При наличие на такава разлика се отбелязва началната честота и се засича до къде продължава тя (`float frequencyStart`, `float frequencyEnd`). В този сектор от честоти се намират ADSR параметрите (атака, разпад, задържане, отпускане).
- при засичане на звук, се създават обекти от модел `Sound` и `SoundOccurance`.

Проектирани са два основни класа – `Sound` и `SoundOccurance`. Класът `Sound` служи като основен модел, който съхранява всеки нов звук засечен от системата (Valchanov, Iliev, 2011), (Valchanov, Petkova, Iliev, 2012).



Фигура 4. Диаграма на класовете съхраняващи извлечените данни

Докато `class Sound` служи като шаблон за физичните данни на всеки разпознат звук, `class SoundOccurance` е създаден, за да съхранява появата на звуци във времето, като запазва тяхното начало и продължителност (Valchanov, Iliev, 2010).

Част 3: Съхранение на извлечената информация

Засечените данни биват съхранени в текстов файл, като в същото време са директно визуализирани в отделен прозорец.

```

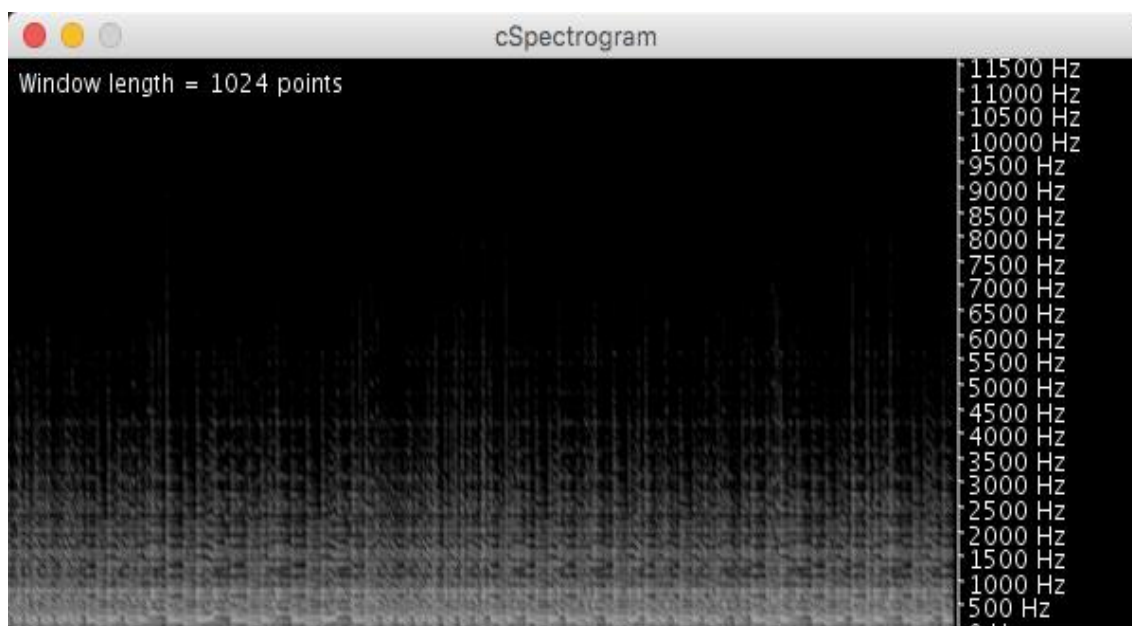
real_time_analysis
The window being used is: Rectangular Window
Time: 246 ms/100

class Sound: FROM: 100hz to 150hz / 4 times. class SoundOccurance: 74.78, 75.27, 164.01, 227.48
class Sound: FROM: 151hz to 236hz / 6 times. class SoundOccurance: 52.95, 53.14, 96.72, 140.48, 183.83, 227.68
class Sound: FROM: 237hz to 279hz / 4 times. class SoundOccurance: 74.95, 75.11, 108.25, 162.08
class Sound: FROM: 280hz to 366hz / 7 times. class SoundOccurance: 31.95, 52.95, 96.72, 118.61, 140.48, 206.18, 227.48
class Sound: FROM: 367hz to 409hz / 2 times. class SoundOccurance: 228.21, 228.41
class Sound: FROM: 410hz to 452hz / 1 times. class SoundOccurance: 227.48
class Sound: FROM: 453hz to 581hz / 1 times. class SoundOccurance: 199.92
class Sound: FROM: 496hz to 538hz / 4 times. class SoundOccurance: 31.95, 52.78, 201.35, 227.48
class Sound: FROM: 539hz to 624hz / 5 times. class SoundOccurance: 52.95, 53.14, 140.11, 172.98, 227.68
class Sound: FROM: 625hz to 839hz / 10 times. class SoundOccurance: 31.95, 53.33, 74.78, 96.36, 118.41, 140.66, 162.08, 183.67,
class Sound: FROM: 668hz to 710hz / 3 times. class SoundOccurance: 162.24, 162.42, 206.75
class Sound: FROM: 754hz to 796hz / 8 times. class SoundOccurance: 53.99, 54.19, 98.59, 119.16, 140.48, 185.81, 208.16, 230.41
class Sound: FROM: 883hz to 925hz / 2 times. class SoundOccurance: 227.68, 227.84
    
```

Фигура 5. Извеждане на получените резултати

Част 4: Визуализация на съхранените данни

За да се създаде визуална репрезентация на спектъра от честоти в реално време е направена спектрограма на звука. С нея допълнително могат да се сравнят крайните данни.



Фигура 6. Спектрограма на входящите данни

6. Заключение и бъдещо развитие

Поради нарастващия интерес към сферата на извличане на информация от звук, полето за изследвания единствено може да расте. Характерът на този вид информация за момента не разрешава създаването на стандарт в обработката. Статията разглежда основите на работата със звук и проблематиката при извличането на символични данни. Представената система предоставя гъвкав подход за справяне на проблема с извличането на информация от звук в реално време. Възможността на системата за вграждане и комбинация на различни методи за работа със звук разрешава използването и в множество сфери на приложение като: системи за разпознаване на музика, извличане на инструменти от звук, създаване на звукови отпечатъци, автоматична транскрипция на музикални изпълнения и др.

Благодарности

Тази работа е подпомогната по проекта МУ17-ФМИ-007 към Фонд „Научни изследвания“ при Пловдивския университет „Паисий Хилендарски“.

Литература

Meurer, W., Aspects of Music Information Retrieval, School of Information at The University of Texas at Austin.

Smith, S., *The Scientist and Engineer's Guide to Digital Signal Processing*, ISBN 0-7506-7444-X.

Wiering, Fr., Can Humans Benefit from Music Information Retrieval?, Department of Information and Computing Sciences, Utrecht University.

Ghias, A., Query By Humming – Musical Information Retrieval in an Audio Database, ACM Multimedia 95 – Electronic Proceedings, November 5–9, 1995, San Francisco, California.

Heinrich A. van Nieuwenhuizen, W. Venter and M. Leenta, The Study and Implementation of Shazam's Audio Fingerprinting Algorithm for Advertisement Identification, Grobler School of Electrical, Electronic and Computer Engineering North-West University, Potchefstroom Campus, South Africa.

Valchanov, N., T. Terzieva, V. Shkurtoov and A. Iliev, Architecture of extensible computation driven systems, Mathematics and mathematical education, *Proc. of 39th spring conference of Union of Bulgarian Mathematicians*, 06–10 April 2010, Albena, Bulgaria, 2010, 207–211.

Iliev, A., G. Hristozov and T. Terzieva, Software environment for dynamic models presentation with statistics, *National Conference "Education in the Information Society"*, Plovdiv, 2006, 38–43.

De Leon, P., Computer Music in Undergraduate Digital Signal Processing, New Mexico State University.

http://www.scottliddell.com/uploaded_images/adsr-725469.png

Processing IDE, <https://www.processing.org/>

Minim Library, <http://code.compartmental.net/minim/>

MidiBus Library, <https://github.com/sparks/themidibus>

Bello, J., L. Daudet, S. Abdallah, C. Duxbury, M. Davies and M. Sandler, *A tutorial on onset detection in music signals*, DOI: 10.1109/TSA.2005.851998

Valchanov, N. and A. Iliev, Implementation of graphical simulation environment for mathematical models, *Fundamental and Complementary Science*, "Mirceacel Batran" Naval Academy Scientific Bulletin, Constanta, Romania, Volume XIV (2), 2011, 222–228.

Valchanov, N., P. Petkova and A. Iliev, Integration of computational library for simulation mathematical models in web-based system for courses management, *Proceedings of the Jubilee National Conference with international participation "Tradition, directions, challenges"*, Paisii Hilendarski University of Plovdiv – branch Smolyan, 50 years Scientific and Educational Institution in the Rhodope Mountains, 19–21.10.2012, 2012, 89–94.

Valchanov, N. and A. Iliev, Workflow Optimization using Intelligent Business Information Systems, *Proceedings of the International Conference "Systems Management Business in Small and Medium Enterprises"*, Svishtov, 23–24 April 2010, 107–112.

Пловдивски университет „П. Хилендарски“,
Факултет по математика и информатика
бул. „България“ № 236, 4003, Пловдив
+35988335303, viktor_matanski@yahoo.com

SYSTEM FOR DATA RETRIEVAL FROM SOUND IN REAL TIME

Viktor Matanski

***Summary:** Retrieving data from sound has attracted a lot of attention for the past 20 years. Research in this area is growing fast and more and more applications are finding their way in our everyday life. From the most obvious application of extracting data from sound and especially music to the more specific extraction data concerning the vibrational phenomenon from moving mechanical parts, human organs, constructions, buildings and infrastructure. Nevertheless, finding patterns is a skill that is being developed by humans from their birth until later life. However, understanding those patterns and turning this skill into proficiency seem to require even more time than that. In this article an idea for a system for information retrieval in real time will be presented. The main accent will be sound data. The aim of the system is to provide an efficient way for the combined use of different algorithms for data retrieval as well as storing the final results as objects that can be further analyzed and used.*

***Keywords:** sound analysis, classification, data retrieval, machine learning*