Parametryzacja sygnałów muzycznych
PO CO?
JAK?

Akustyka Muzyczna

Dane

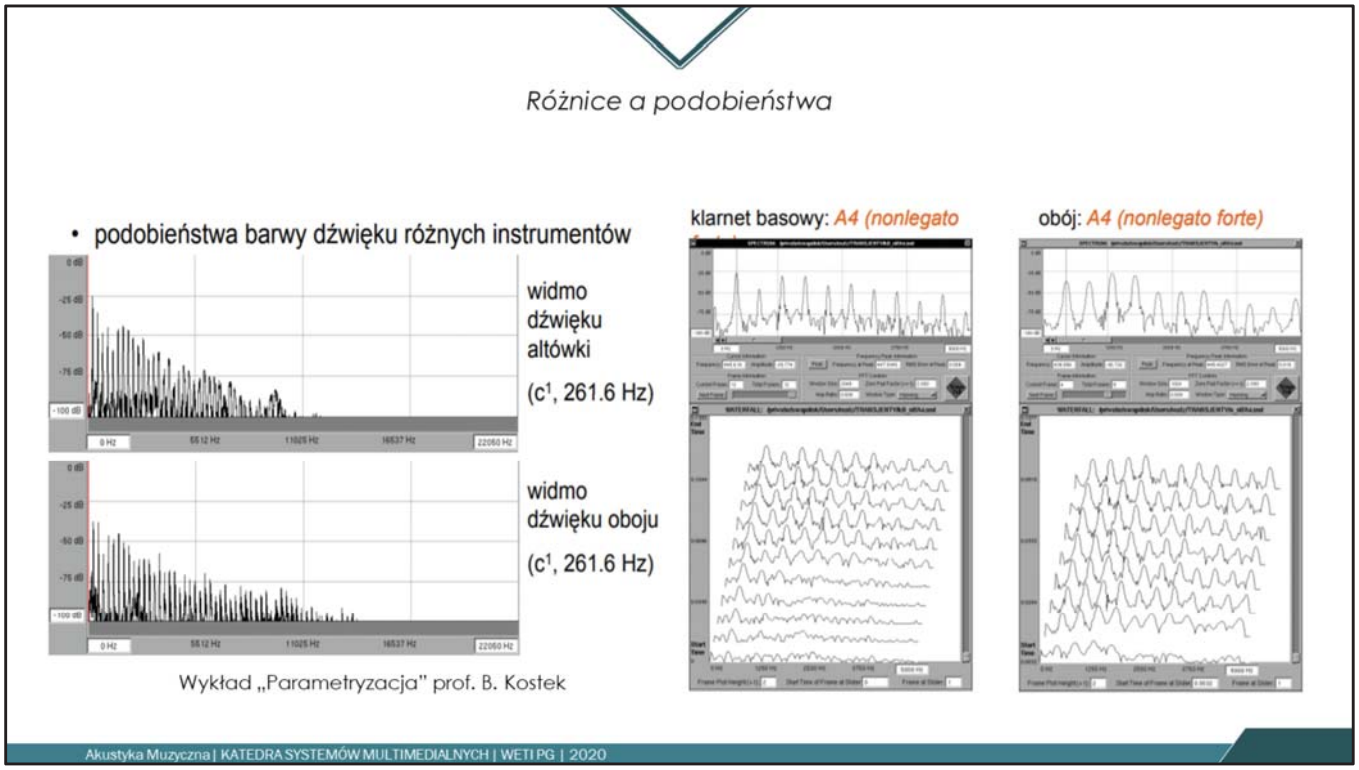
- Hipoteza – Możliwe jest automatyczne rozróżnianie:
  - instrumentów muzycznych,
  - gatunków muzycznych,
  - nastroju w utworze,
  - separacja ścieżek w utworze,
  - Rozróżnianie dźwięków różnych instrumentów
  - …
- Co jestem w stanie wyznaczyć na podstawie posiadanej bazy danych?
  - Wyszukiwanie podobieństw w zbiorze parametrów (fishing)
  - Wyznaczenie konkretnych cech charakterystycznych dla danych obiektów (dźwięków)

Reprezentacja danych

Data Science:

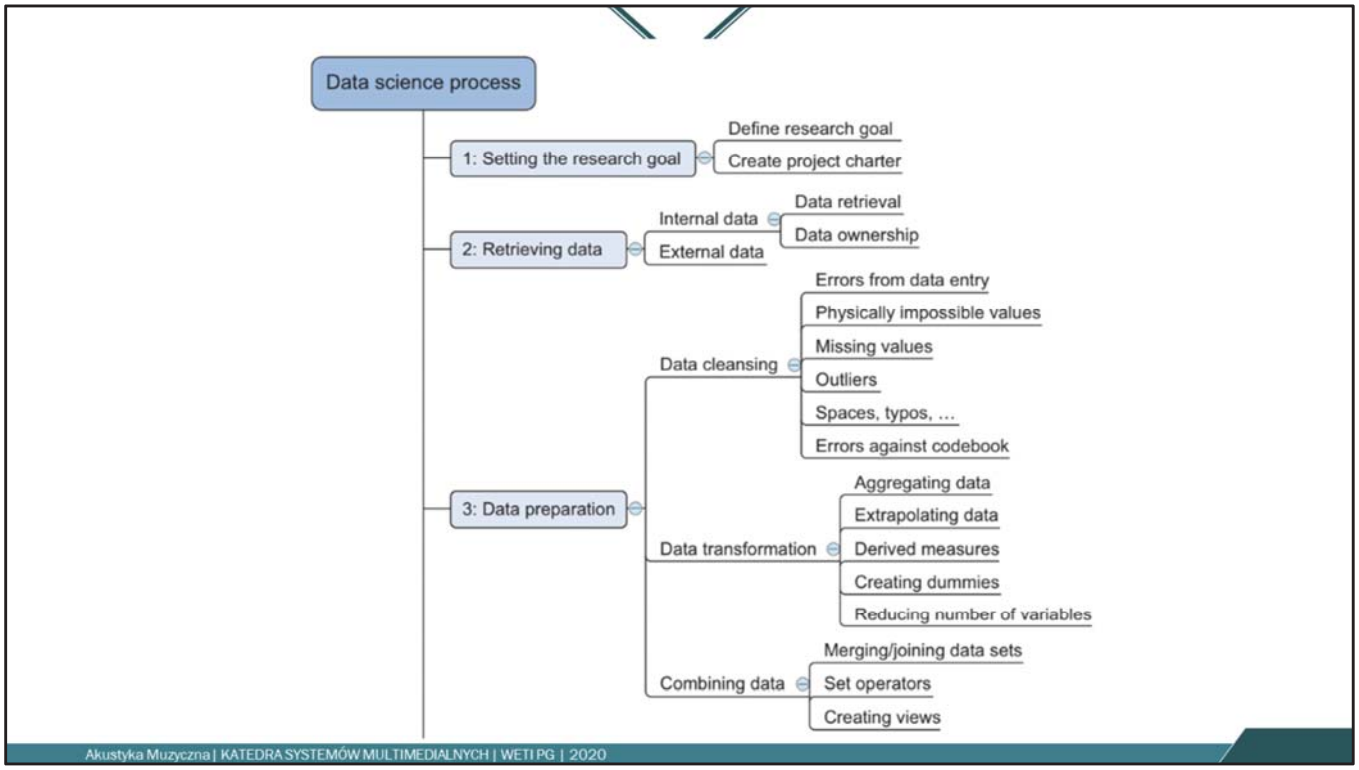
- Strukturalne – typowo w postaci relacyjnej bazy danych, tabeli, skoroszytów.
- A-strukturalne (trudne do umieszczenia w bazie, audio, obrazy, bloki tekstu)
- Ilościowe
- Jakościowe
- Big data

- Pliki audio w formatach muzycznych
- Deskryptory (np. MPEG-7),
- Liczbowe miary opisu dźwięku:
  - czasowe,
  - częstotliwościowe,
  - częstotliwościowo czasowe,
- Obrazowe miary opisy dźwięku:
  - waveform,
  - Spektrogram (mel-, bark-… różne odmiany),
  - Gramathone,
  - …



Przy analizowaniu dźwięków różnych instrumentów, część parametrów będzie taka sama, lub podobna. Kluczem parametryzacji jest znalezienie takich parametrów, które z przyjętą przez nas dokładnością rozróżnią dwa różne brzmienia.

Ogólny schemat działania systemu rozpoznania muzyki można przedstawić za pomocą tego schematu blokowego.
W zależności od analizowanego systemu, dane mogą być rejestrowane w sposób ciągły, wymuszony (push to record), lub jako nagrania ciągiem w wybranych środowiskach (las do pozyskania odgłosów konkretnych zwierząt).
Proces przetwarzania wstępnego, w zależności od typu sygnału może dotyczyć innych spraw technicznych. W systemach typu push to record, przetwarzanie wstępne może dotyczyć wstępnej analizie sygnałowej zarówno w dziedzinie czasu, częstotliwości jak i ich kombinacji. Mogą być wprowadzany procesy normalizacji sygnału, czy też odszumiania. Przy bazach tworzonych z pełnych nagrań, proces ten dotyczy głównie pozyskiwania próbek – fragmentów z nagrań zawierających obiekt zainteresowania. Może się to odbywać ręcznie, lub z pomocą wstępnych algorytmów detekcji.
Proces ekstrakcji cech będzie polegał na wyznaczeniu zadanych cech obiektu wg. zadanych schematów (MPEG-7, inne parametry typu MIR).
Kolejny etap jest opcjonalny, zależy od typu systemu.
Na końcu mamy klasyfikacje obiektu wg. wybranej metody.



Wstępna analiza danych

Liczba danych

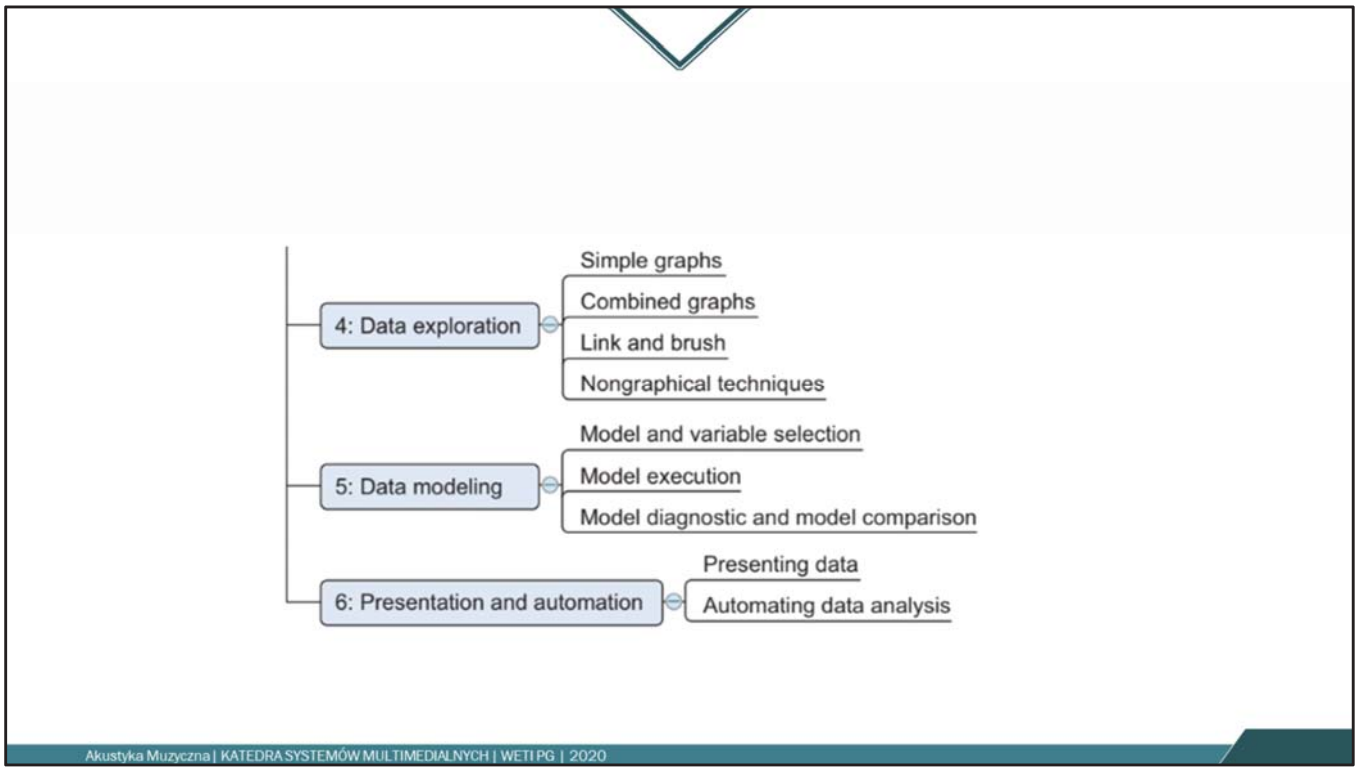
- Ile zmiennych (cech obiektu)
- Ile przypadków (obiektów)

Typy danych

- Dane jakościowe (opisowe)
- Dane ilościowe (liczbowe)

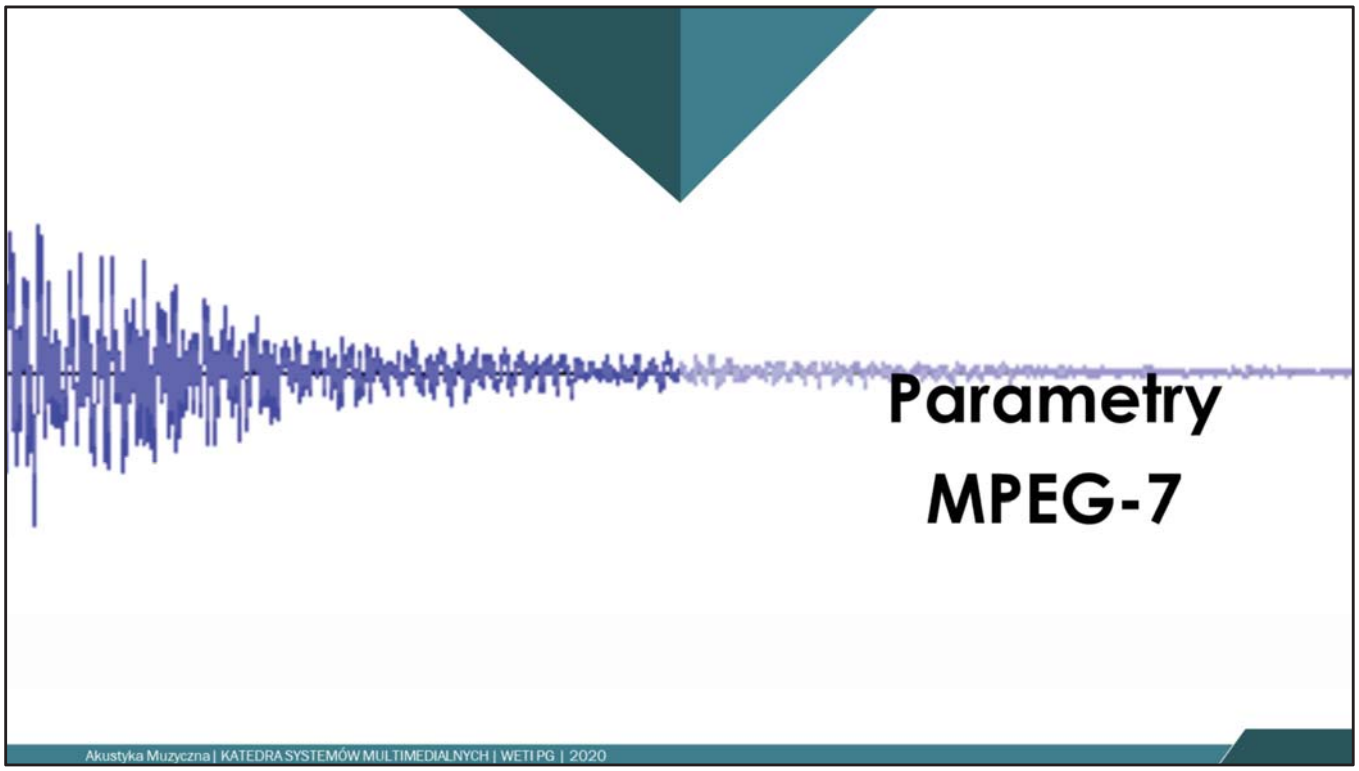
Niepełne dane

DAVY CIELEN, ARNO D. B. MEYSMAN, MOHAMED ALI: Introducing Data Science BIG DATA, MACHINE LEARNING, AND MORE, USING PYTHON TOOLS



Introducing
Data Science
BIG DATA, MACHINE LEARNING,
AND MORE, USING PYTHON TOOLS

DAVY CIELEN
ARNO D. B. MEYSMAN
MOHAMED ALI


Parametry
MPEG-7
Akustyka Muzyczna | KATEDRA SYSTEMÓW MULTIMEDIALNYCH | WETI PG | 2020

Multimedia Content Description Interface

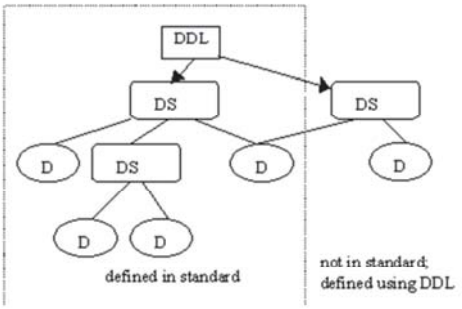
MPEG-7 → "Moving Pictures Expert Group" ISO/IEC JTC1/SC29/WG11

Standard wykorzystujący XLM (*i XML Schema*) do przechowywania danych dotyczących plików multimedialnych (z zastosowaniem timecode).

MPEG-7 standaryzuje:

- schematy deskrypcji i same deskryptory (Descriptor, D)
- język opisu (Description Definition Language, DDL),
- schemat kodowania deskrypcji (Description Scheme, DS).

Pierwsza oficjalna dokumentacja datowana jest na 2002 rok.



scope of MPEG-7

Akustyka Muzyczna | KATEDRA SYSTEMÓW MULTIMEDIALNYCH | WETI PG | 2020

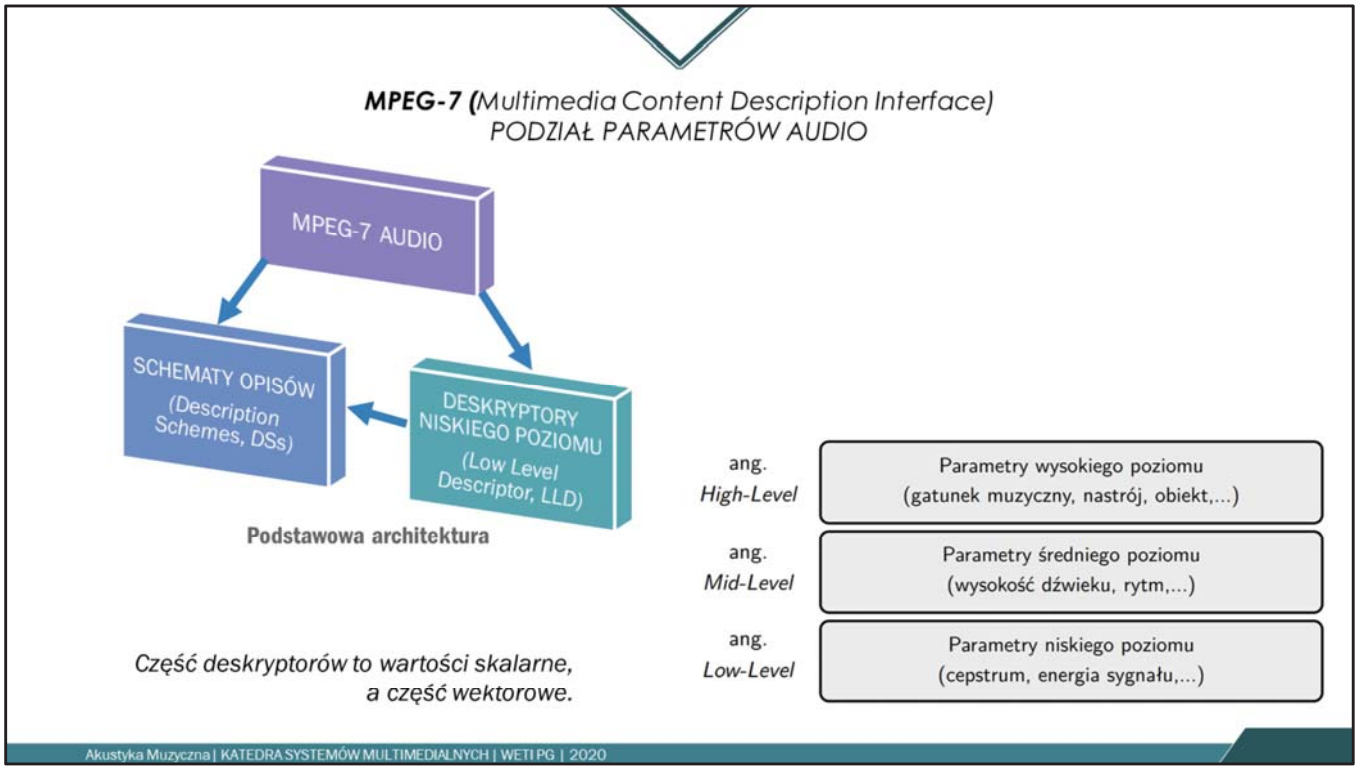
Standard MPEG7 jest w pełni udokumentowany. Do ważniejszych dokumentów dotyczących audio należy wymienić:

ISO/IEC 15938-4:2002 Information technology — Multimedia content description interface — Part 4: Audio

ISO/IEC TR 15938-8:2002 Information technology — Multimedia content description interface — Part 8: Extraction and use of MPEG-7 descriptions

Dostęp do norm jest płatny. Można uzupełnić wiedzę za pomocą innych dostępnych źródeł (http://mpeg7.doc.gold.ac.uk/).

http://www.cs.bilkent.edu.tr/~bilmdg/bilaudio-7/MPEG7.html

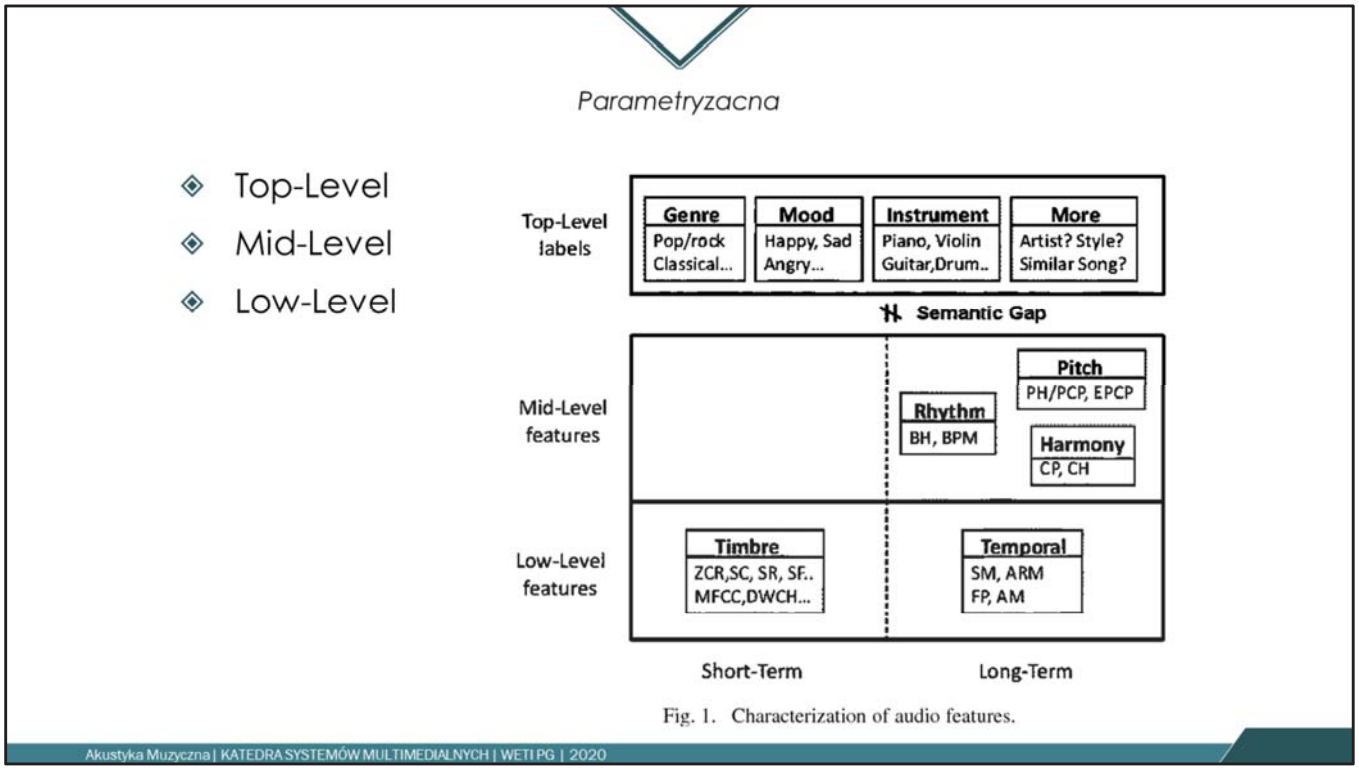


Standard MPEG7 jest w pełni udokumentowany. Do ważniejszych dokumentów dotyczących audio należy wymienić:

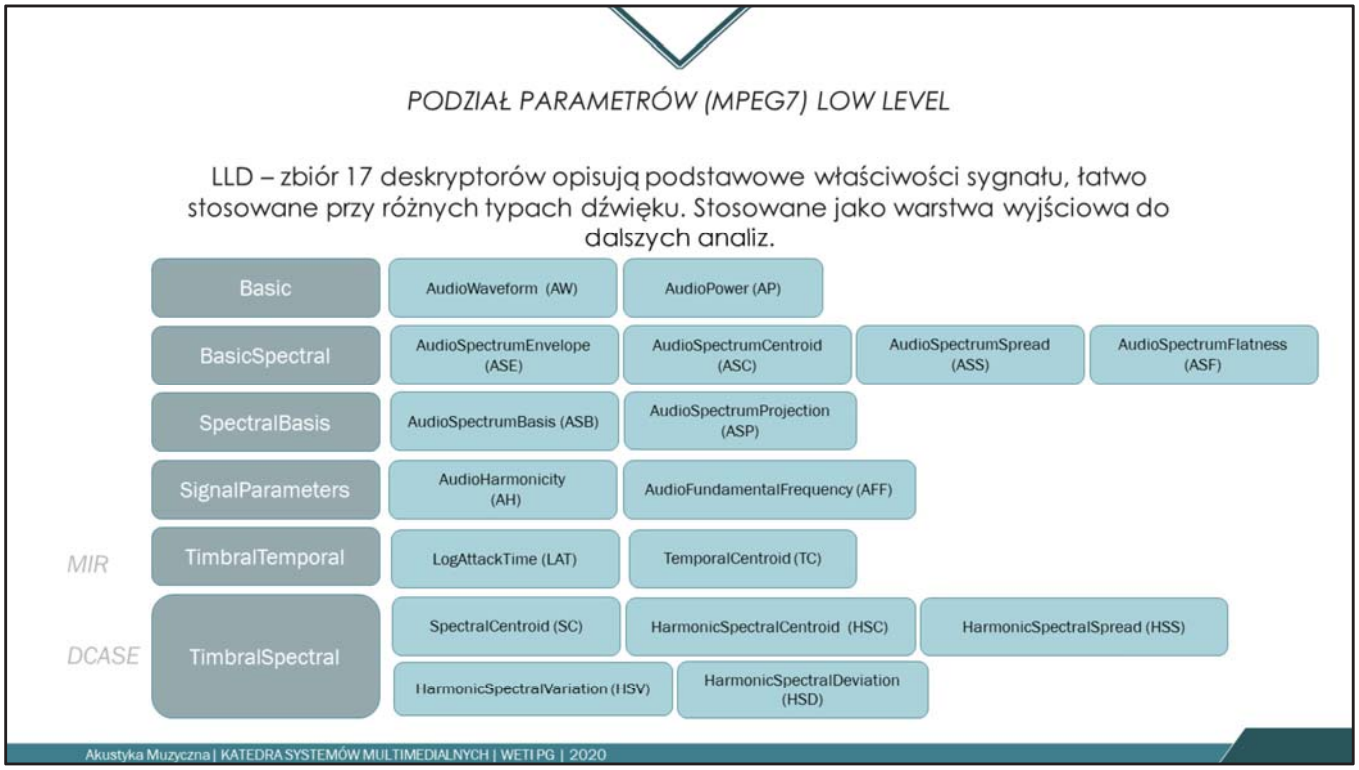
ISO/IEC 15938-4:2002 Information technology — Multimedia content description interface — Part 4: Audio

ISO/IEC TR 15938-8:2002 Information technology — Multimedia content description interface — Part 8: Extraction and use of MPEG-7 descriptions

Dostęp do norm jest płatny. Można uzupełnić wiedzę za pomocą innych dostępnych źródeł (http://mpeg7.doc.gold.ac.uk/).



https://www.analyticsvidhya.com/blog/2018/01/10-audio-processing-projects-applications/



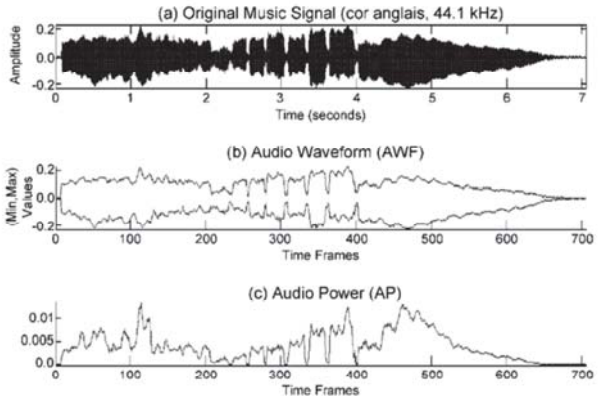
Grupa podstawowa (Basic) zawiera głównie szybki i prosty opis dotyczący głownie kształtu waveformu. Celem AW jest głownie wizualizacja sygnału w edytorze poprzez zobrazowania minimalnych i maksymalnych wartości sygnału/ramki sygnału. AP określa energie sygnału w czasie (ramka) i wyznaczane jest jako średnia kwadratowa.
BasicSpectral – grupa parametrów opisujących podstawowe właściwości pasma sygnału. ASE to krótkookresowy opis energii widma w pasmach sygnału w skali logarytmicznej. Pasmo ograniczone jest zakresem słyszalności dla ludzkiego narządu słuchu. ASC wyznacza środek ciężkości widma wyznaczonego parametrem ASE. Wartość ASC informuje nas czy mamy do czynienia z dźwiękiem jasnym, czy ciemnym. ASS przedstawia z kolei wariancje energii widma sygnału od środka ciężkości, pozwala na separacje dźwięków tonalnych od szumowych. ASF – opisuje jak bardzo obwiednia sygnału (energetyczna częstotliwościowa) odbiega od płaskiego układu. Jest to kolejny parametr pozwalający na separacje dźwięków tonalnych od szumowych.
SignalParameters – to grupa kolejnych raczej bazowych parametrów dźwięku: AFF – częstotliwość podstawowa dźwięku, AH wyznacza stopnień harmoniczności sygnału. Wyznaczany jest w oparciu o stosunek składowych harmonicznych do pozostałych w sygnale (*harmonic ratio)* i *upper limit od harmonicity*. W przypadku czystego, harmonicznego sygnału wartość AH = 1, przy dźwięku nieposiadającym znacząco harmonicznych składowych AH = 0.
SpectralBasis – stosowane głownie przy rozpoznawaniu dźwięku. ASB przekształca widmo sygnału ograniczając jego wymiarowość w oparciu o statystykę. ASP działa podobnie jak ASB, jednak sygnał analizowany jest w skali decybelowej.
TimbralTemporal – typowo stosowane w systemach MIR. LAT to logarytm czasu ataku dźwięku (obwiednia ADSR), dzięki czemu jesteśmy w stanie ocenić czy dźwięk jest gwałtowny, czy raczej stonowany. TC opisuje chwile w której skupiona jest energia sygnału.
TimbralSpectral - deskryptory bazują na estymacji harmonicznych sygnału. HSC to amplitudowo ważona średnia harmonicznych w sygnale. HSS – zamiast średniej wyznacza odchylenie. HSD – średnia harmonicznych z obiedniej częstotliwościowej sygnału, uwzględnia wartości sąsiadujących harmonicznych. HSV – korelacja harmonicznych. S.C. – średnia ważona (energią) częstotliwości w sygnale. Grupa tych deskryptorów stosowana jest głownie w rozpoznawaniu dźwięków środowiskowych.

PODZIAŁ PARAMETRÓW (MPEG7) LOW LEVEL

Basic | AudioWaveform (AW) | AudioPower (AP)

$$AP(l) = \frac{1}{N_{hop}} \sum_{n=0}^{N_{hop}-1} \left| s(n + lN_{hop}) n \right|^2, (0 \le l \le L-1)$$

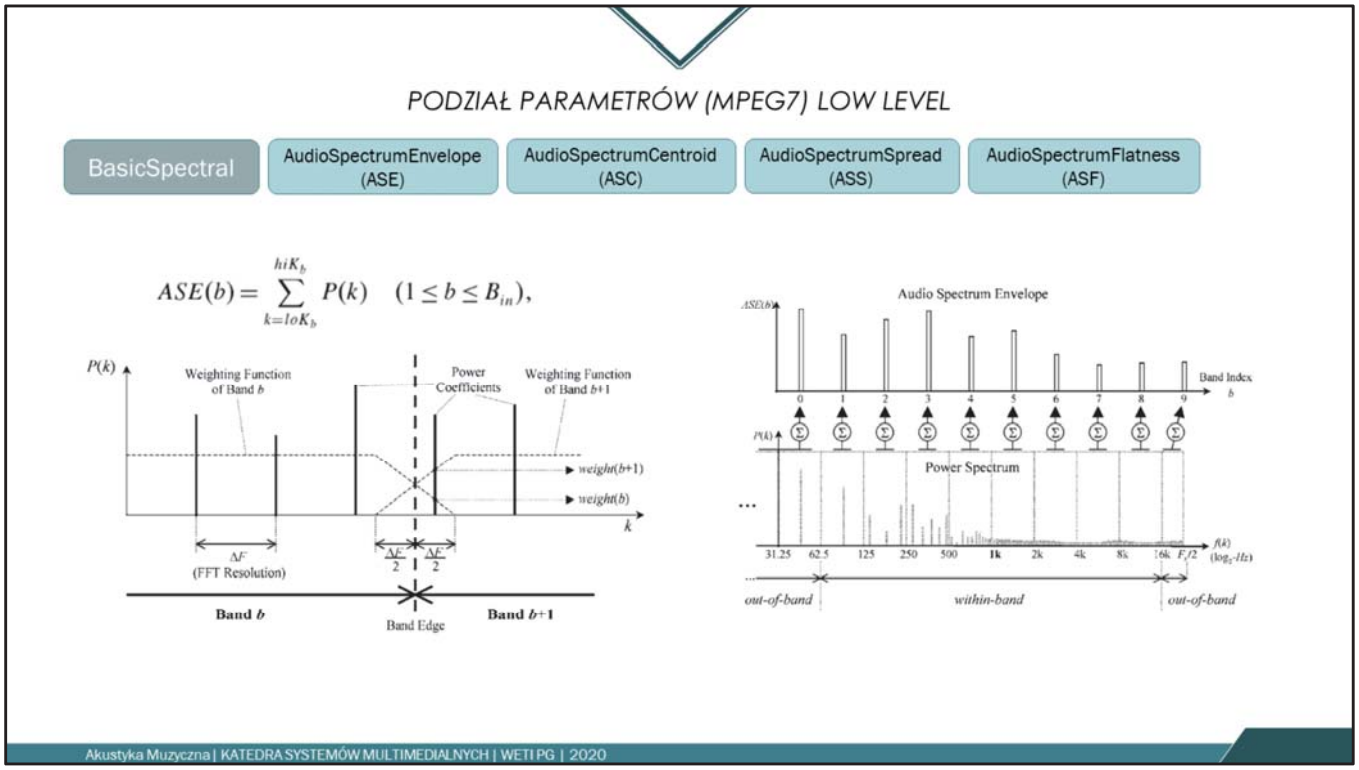
Gdzie L – całkowita liczba ramek



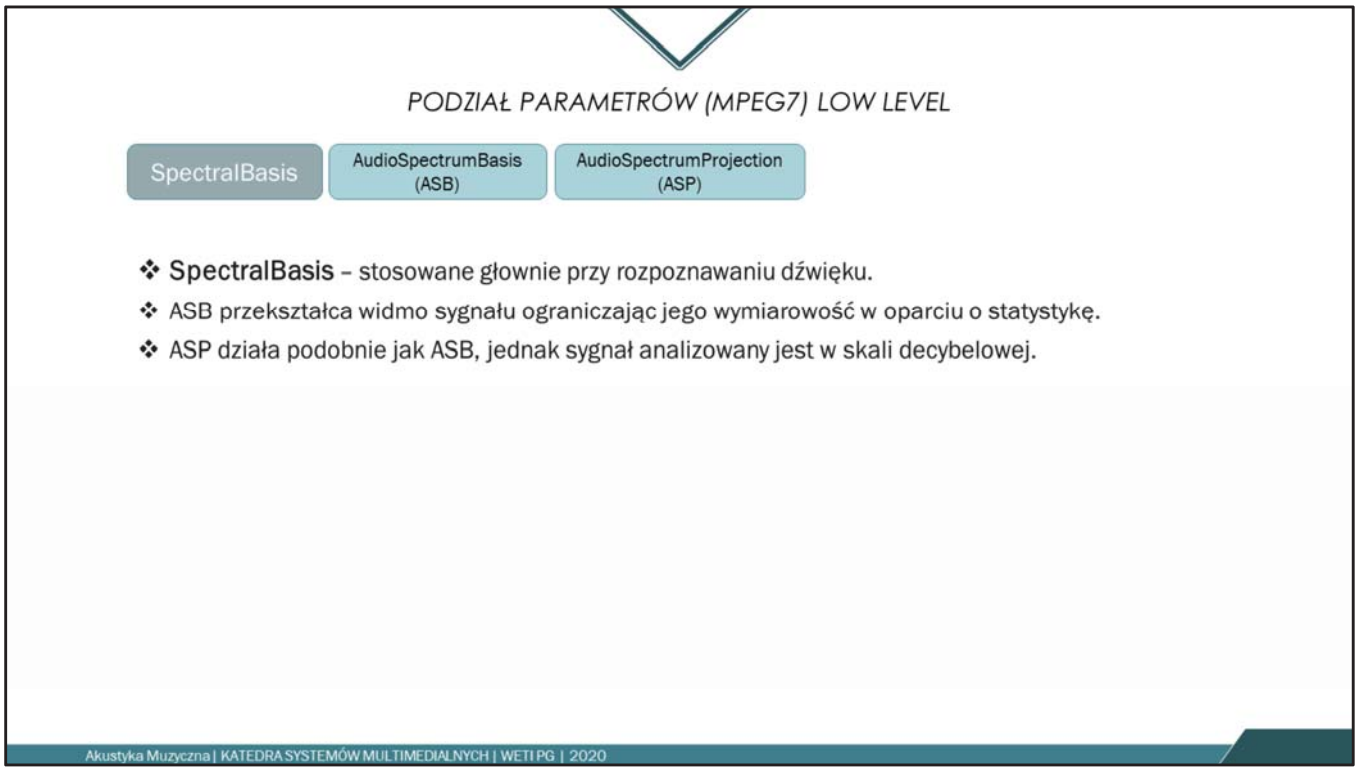
MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval
By Hyoung-Gook Kim, Nicolas Moreau, Thomas Sikora

Grupa podstawowa (Basic) zawiera głównie szybki i prosty opis dotyczący głownie kształtu waveformu. Celem AW jest głownie wizualizacja sygnału w edytorze poprzez zobrazowania minimalnych (minRange) i maksymalnych(maxRange) wartości sygnału/ramki sygnału. AP określa energie sygnału w czasie (ramka) i wyznaczane jest jako średnia kwadratowa.

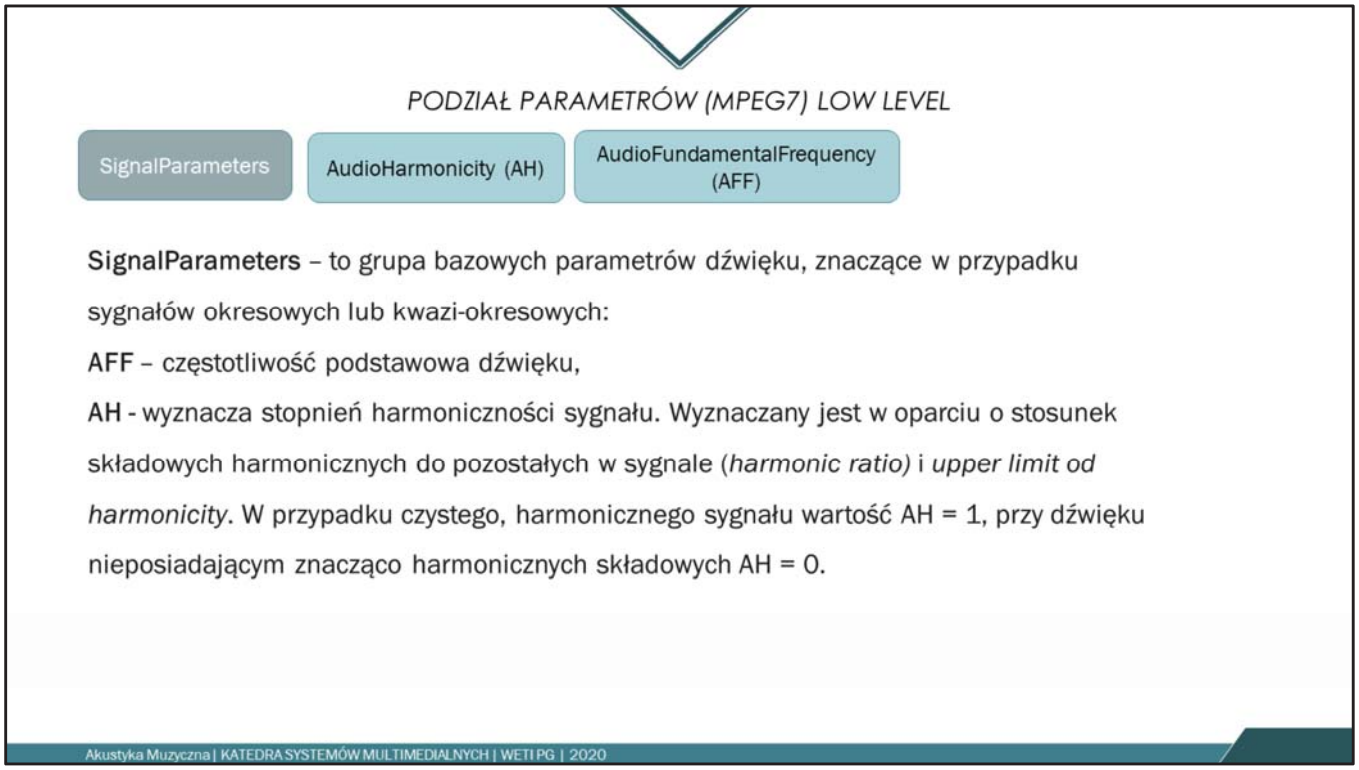
$$AP(l) = \frac{1}{}$$



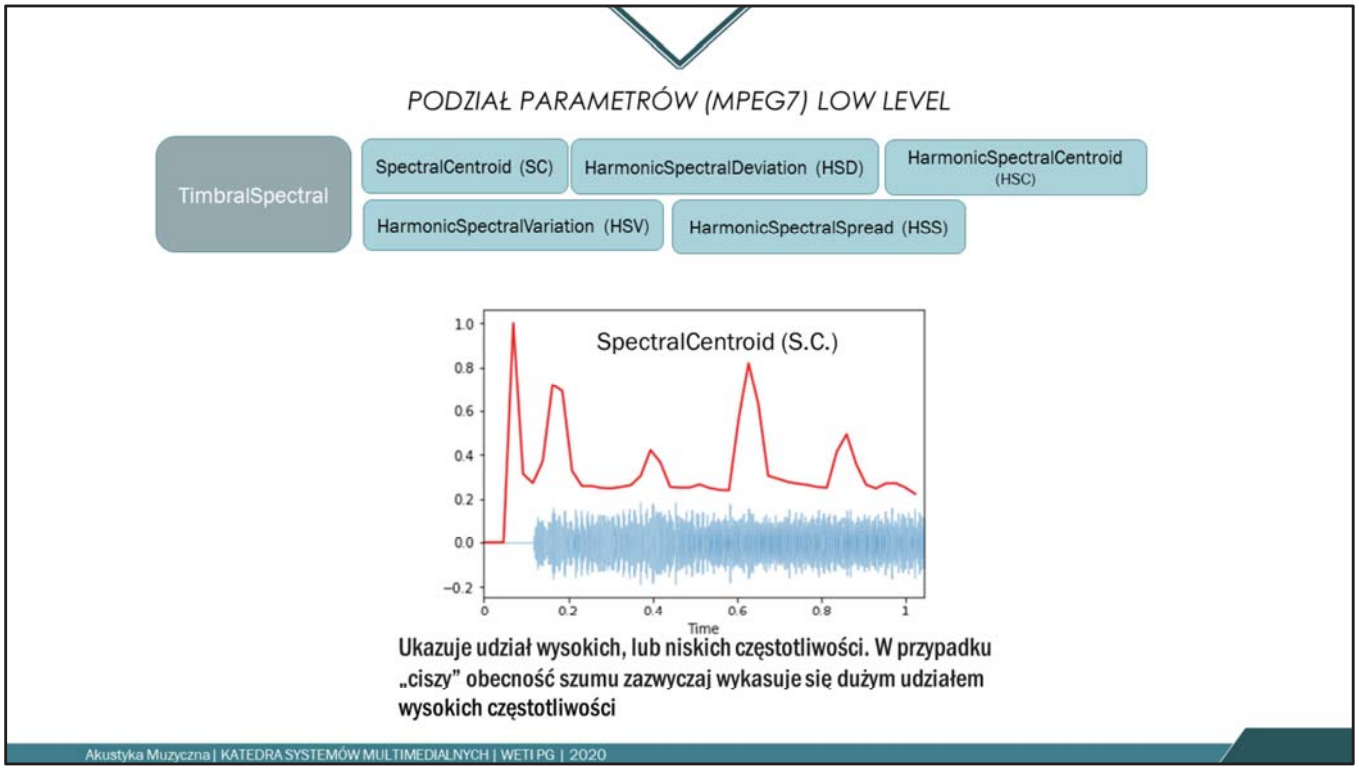
BasicSpectral – grupa parametrów opisujących podstawowe właściwości pasma sygnału.
ASE to krótkookresowy opis energii widma w pasmach sygnału w skali logarytmicznej. Pasmo ograniczone jest zakresem słyszalności dla ludzkiego narządu słuchu, przyjęto zakres od 62,5 do 16000Hz. Ze względu na ten podział, w obliczeniach uzyskujemy wartości średnie i wariancje dla podpasm(ASE1-34, ASRv1-v34), jak i uśrednioną wartość ASE_M, ASE_MV.
ASC wyznacza środek ciężkości widma wyznaczonego parametrem ASE. Wartość uzyskana oznacza odległość od częstotliwości referencyjnej (1kHz) w oktawach. Wartość ASC informuje nas czy mamy do czynienia z dźwiękiem jasnym, czy ciemnym.
ASS przedstawia z kolei wariancje energii widma sygnału od środka ciężkości, pozwala na separacje dźwięków tonalnych od szumowych. Wynikiem jest wartość średnia i wariancja (ASS, ASS_V).
ASF – opisuje jak bardzo obwiednia sygnału (energetyczna częstotliwościowa) odbiega od płaskiego układu. Jest to kolejny parametr pozwalający na separacje dźwięków tonalnych od szumowych. Wyznaczany jest ze stosunku współczynników widma średniej geometrycznej i arytmetycznej.



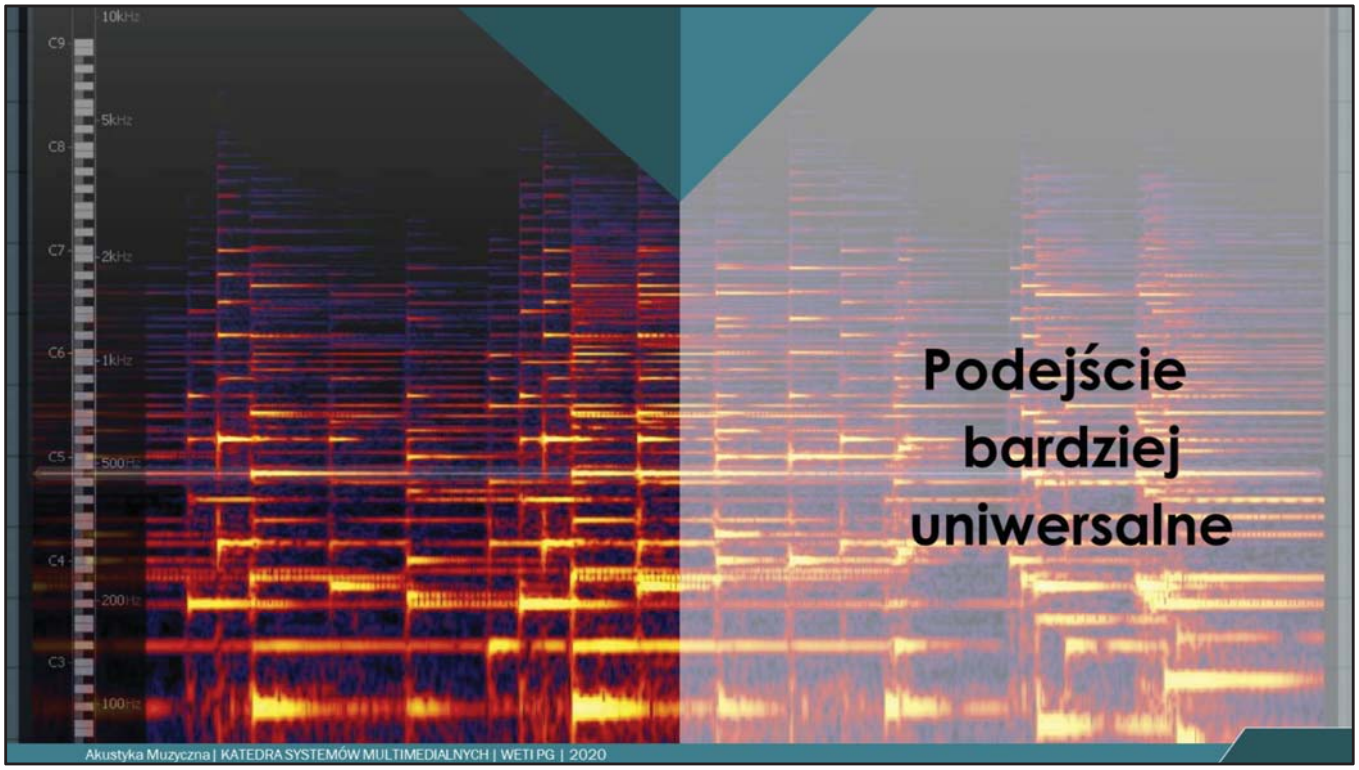
SpectralBasis – stosowane głownie przy rozpoznawaniu dźwięku.
ASB przekształca widmo sygnału ograniczając jego wymiarowość w oparciu o statystykę. Jest w pewnym sensie projekcją wielowymiarowego opisu w mniej wymiarową reprezentacje. Przedstawia stasytyke sygnału dla poszczególnych segmentów dźwięku.
ASP działa podobnie jak ASB, jednak sygnał analizowany jest w skali decybelowej.



SignalParameters – to grupa kolejnych raczej bazowych parametrów dźwięku, znaczące w przypadku sygnałów okresowych lub kwazi-okresowych:
AFF – częstotliwość podstawowa dźwięku,
AH wyznacza stopnień harmoniczności sygnału. Wyznaczany jest w oparciu o stosunek składowych harmonicznych do pozostałych w sygnale (*harmonic ratio)* i *upper limit od harmonicity*. W przypadku czystego, harmonicznego sygnału wartość AH = 1, przy dźwięku nieposiadającym znacząco harmonicznych składowych AH = 0.
Harmoniczne dźwięki do muzyka i mowa, nieharmoniczne to szum lub hałas (aharmoniczny, złożony z kilkunastu źródeł).



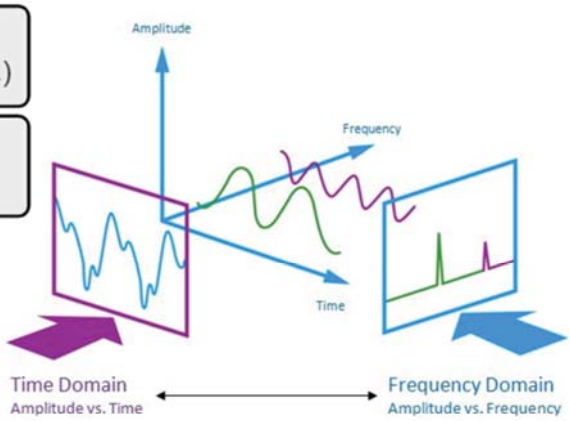
TimbalSpectral - deskryptory bazują na estymacji harmonicznych sygnału.
HSC to amplitudowo ważona średnia harmonicznych w sygnale.
HSS – zamiast średniej wyznacza odchylenie (HSS, HSS_V).
HSD – średnia harmonicznych z obiedniej częstotliwościowej sygnału, uwzględnia wartości sąsiadujących harmonicznych.
HSV – korelacja harmonicznych. Wyznaczana jako wartość średnia w czasie (HSV) i wariancja (HSV_V).
S.C. – średnia ważona (energią) częstotliwości w sygnale. Grupa tych deskryptorów stosowana jest głownie w rozpoznawaniu dźwięków środowiskowych.


10kHz
C9
5kHz
C8
C7
2kHz
C6
1kHz
C5
500Hz
C4
200Hz
C3
100Hz
Podejście
bardziej
uniwersalne
Akustyka Muzyczna | KATEDRA SYSTEMÓW MULTIMEDIALNYCH | WETI PG | 2020

PODZIAŁ PARAMETRÓW (dziedzina)

ang. *Time-Frequency*	Parametry czasowo-częstotliwościowe (spektrogram, wartości średnich, odchylenia standardowego i momentów statystycznych,...)
ang. *Frequency*	Parametry częstotliwościowe (f0, Power Spectral Density (PSD), harmoniczność,...)
ang. *Time*	Parametry czasowe (zcr, energia sygnału, obwiednia,...)

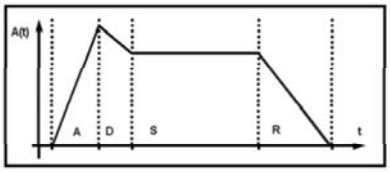
W zależności od typu parametru i podejścia do analizy możemy zbudować wektor cech również o parametry statystyczne dotyczące zmian sygnału w czasie i/lub częstotliwości.



https://towardsdatascience.com/extract-features-of-music-75a3f9bc265d

PODZIAŁ PARAMETRÓW (dziedzina)

- Obwiednia ADSR (Attack-Decay-Sustain-Relese) – czas trwania poszczególnych etapów
- Parametry Tristimulus
- MFCC
- Zero Crossing



$$Tr_1(t) = A_1^2(t) \Big/ \sum_{n=1}^{N} A_n^2(t)$$

$$Tr_2(t) = \sum_{n=2}^{4} A_n^2(t) \Big/ \sum_{n=1}^{N} A_n^2(t)$$

$$Tr_3(t) = \sum_{n=5}^{n} A_n^2(t) \Big/ \sum_{n=1}^{N} A_n^2(t)$$

A(i) – amplituda *i*-tej składowej,
n - ilość próbek sygnału

Przykładowym parametrem w rozpoznawaniu dźwięków muzycznych jest np. dobrze nam znana obwiednia ADSR. W procesie rozpoznawania dźwięków, w większości instrumentów, największe znaczenie ma czas ataku, narastania. Dla instrumentów szarpanych z kolei mówimy o fazie wybrzmiewania (nie ma tu też fazy ustalonej). Parametry Tristimulus – rozróżnianie dźwięków uzyskuje się na podstawie analizy zawartości grup harmonicznych widma względem całkowitej sumy amplitud harmonicznych.

Fastl, Zwicker

- Parametry psychoakustyczne(Fastl, Zwicker)
- Stosowane częściej jako badanie jakości instrumentu niż na potrzeby rozpoznawania dźwięków
- Główne parametry:
  - Loudness
  - Sharpness, Pleasantness
  - Fluctuation Strength
  - Roughness
  - Rhythm

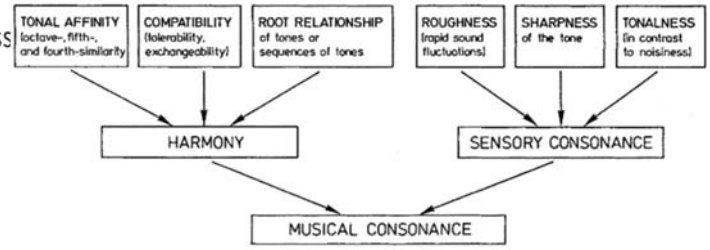


Fig. 16.47. Concept of musical consonance

Typowo parametry te stosowane są w badaniach uciążliwości urządzeń (odkurzacze, silniki itp.). Jednak można z nich skorzystać przy dźwiękach z natury przyjemniejszych. Jednak, częściej spotyka się nie przy badaniu efektywności i właściwości instrumentów muzycznych i części z jakich są wykonane – jakość (albo zdolność) strun, ustników wykonanych z różnych materiałów itp..

Matlab - MIR

- Łatwa powiązana miedzy sobą biblioteka działająca w środowisku matlab z narzędziami do klasyfikacji
- Podział parametrów na:
  - Dynamika (rms, lowenergy,...)
  - Rhythm (fluctuaction, beatspectrum, tempo...)
  - Timbre (atacktime, zerocross, roll-off...)
  - Pitch (pitch, midi)
  - Tonality (chromagram, tonalcentroid, ...)

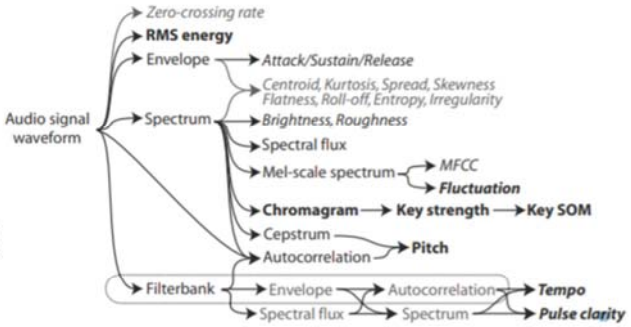


Fig. 1. Overview of the musical features that can be extracted with *MIRToolbox*.

Najpopularniejszy pakiet parametrów w środowisku matlab
Dynamika, tyrm, barwa
https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox/manual1-7-2.pdf

https://link.springer.com/chapter/10.1007/978-3-540-78246-9_31

python

- Librosa
  - Spectral features (chroma, melspectrogram, rms, mfcc...)
  - Rhythm (tempogram)
- pyAudio Analysis

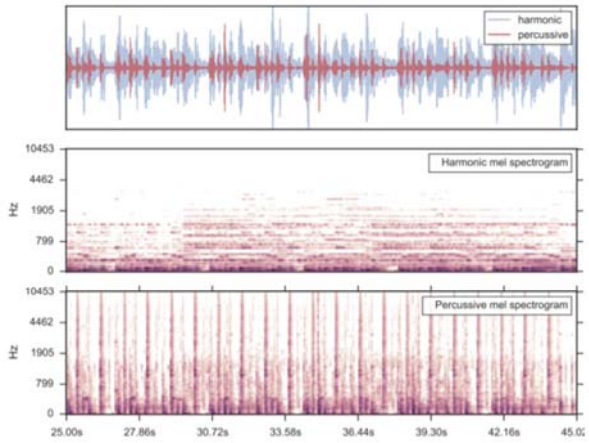


Fig. 4: *Top: the separated harmonic and percussive waveform Middle: the Mel spectrogram of the harmonic component. Botton the Mel spectrogram of the percussive component.*

https://librosa.github.io/librosa/
http://conference.scipy.org/proceedings/scipy2015/pdfs/brian_mcfee.pdf
https://towardsdatascience.com/extract-features-of-music-75a3f9bc265d

https://github.com/tyiannak/pyAudioAnalysis
https://www.kdnuggets.com/2020/02/audio-data-analysis-deep-learning-python-part-1.html
https://github.com/tyiannak/multimodalAnalysis

http://marsyas.info/downloads/datasets.html


Obszary zastosowań
AMAZON.COM/CNNMONEY
Akustyka Muzyczna | KATEDRA SYSTEMÓW MULTIMEDIALNYCH | WETI PG | 2020

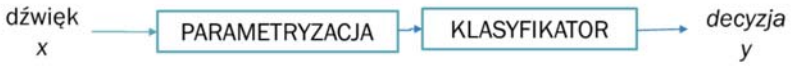
OBSZARY AI W AUDIO

1. Klasyfikacja (*Audio Classification*)
2. Detekcja początków (*Onset detection*)
3. Fingerprinting (*Audio Fingerprinting*)
4. Automatyczne tagowanie (*Automatic Music Tagging*)
5. Segmentacja (*Audio Segmentation*)
6. Separacja źródeł (*Audio Source Separation*)
7. Śledzenie rytmu (*Beat Tracking*)
8. Systemy rekomendacji (*Music Recomendation*)
9. Pozyskiwanie informacji o muzyce (*Music Information Retrieval*)
10. Transkrypcja (*Music Transcription*)

1

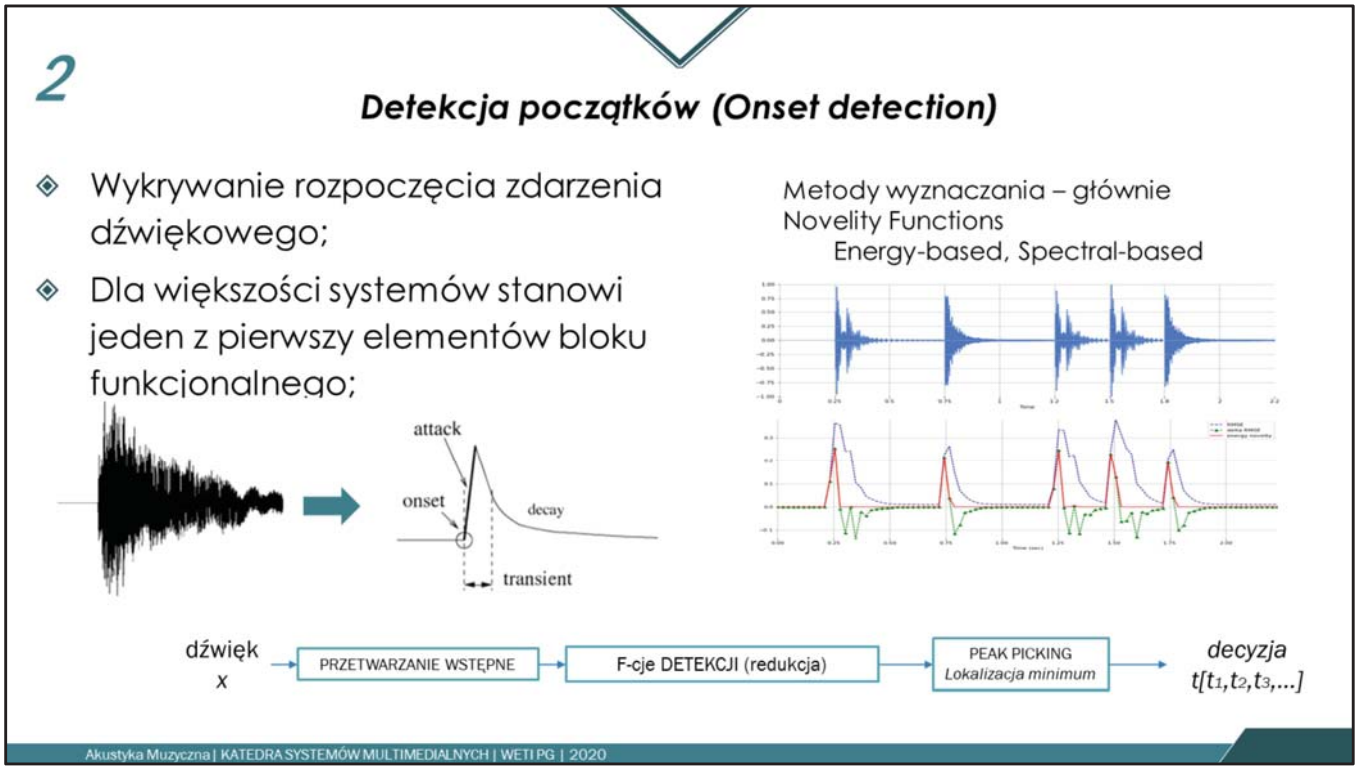
Klasyfikacja (Audio Classification)

- Podstawowy problem MH (Machine Hearing)

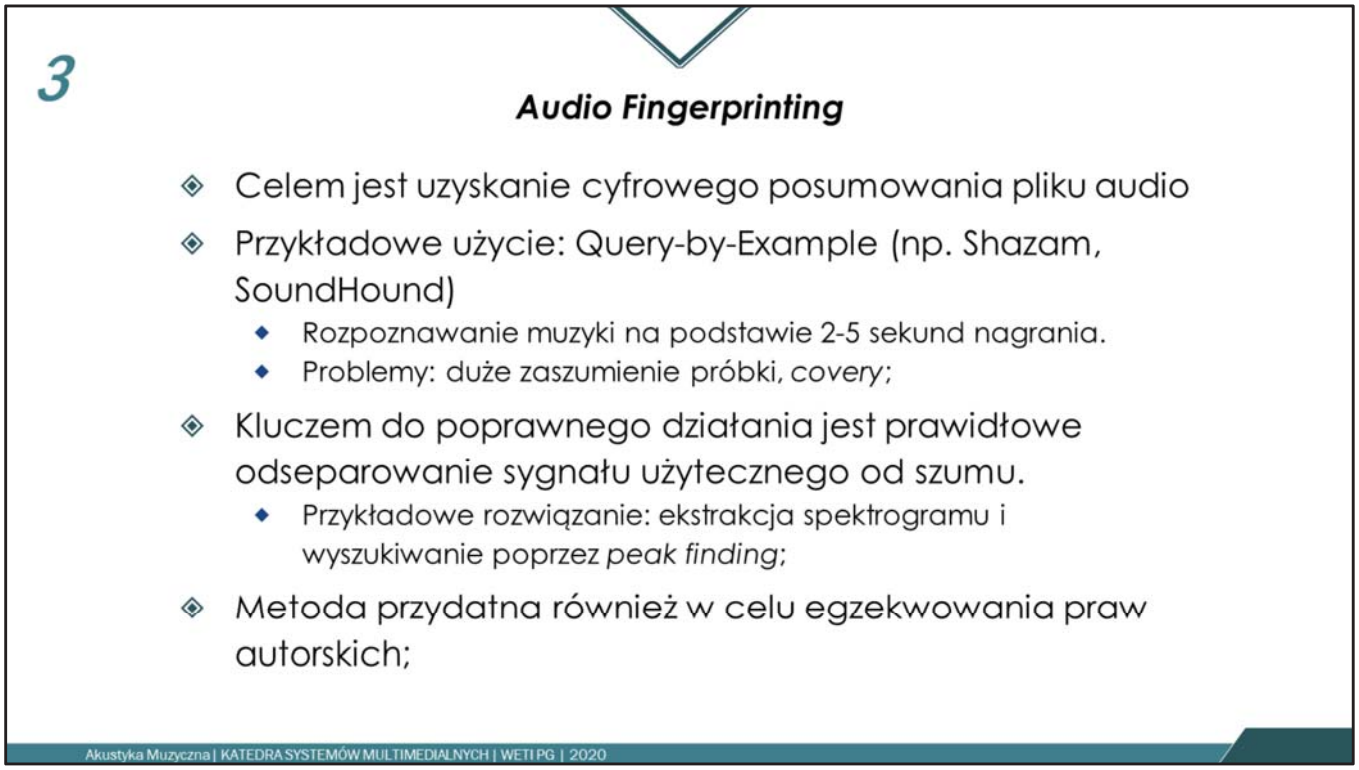


- Typowe problemy klasyfikacji:
  - gatunek muzyczny,
  - instrument muzyczny,
  - nastrój,
  - wykonawca,
  - (klasyfikacja zdarzeń, lokalizacji nagrań w DCASE)

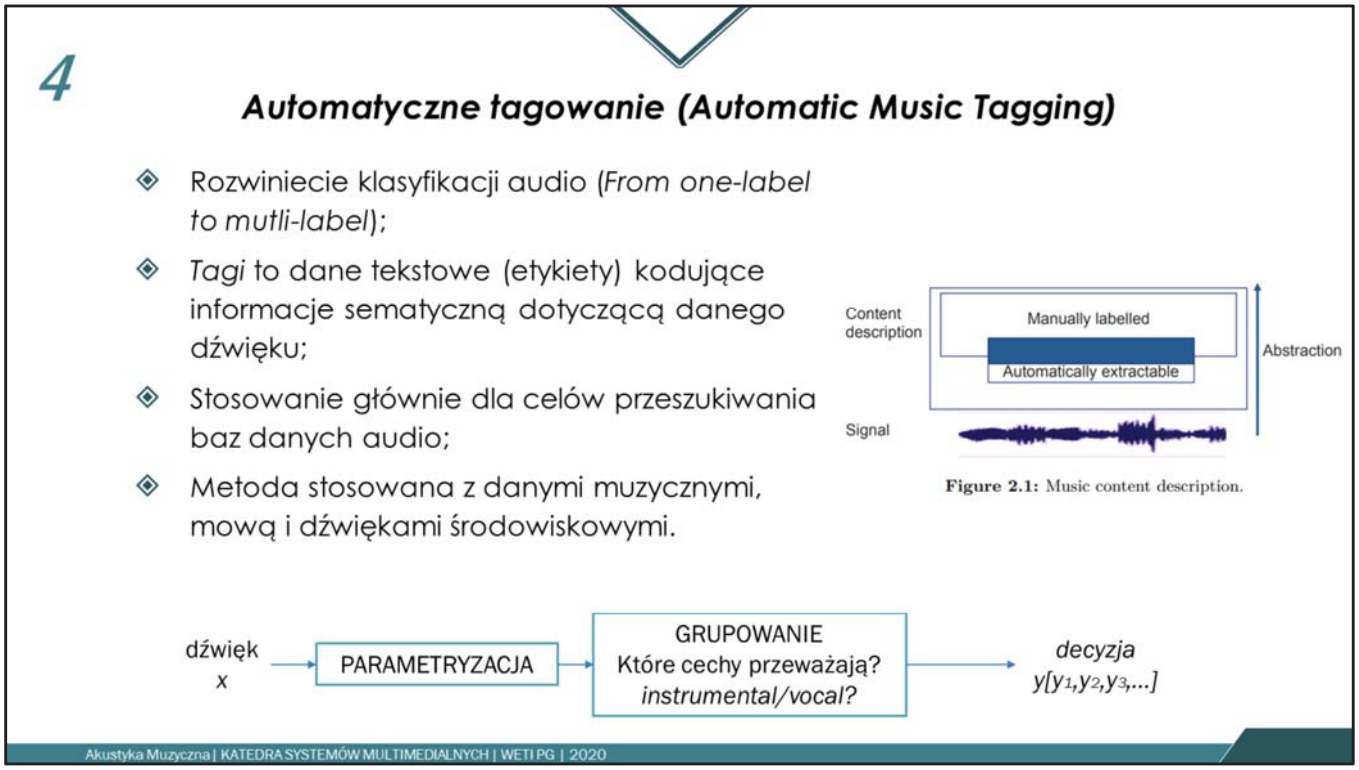
https://ieeexplore.ieee.org/document/5664796?reload=true



Wykrywanie rozpoczęcia sygnału stosowane jest w kilku dziedzinach, nie tylko związanych z muzyką. Dotyczy również sygnałow biologicznych (EKG), danych środowiskowych (sejsmogram), jak i przy analizie zachowań giełdowych. Użyte narzędzia zależą od charakteru sygnału zmiennego w czasie.
Cały proces podobnie jak w innych obszarach ML w audio można podzielić na trzy etapy. Przy przetwarzaniu wstępnym można zastosować narzędzia ułatwiające zadziałanie funkcji detekcji. Przykładowo separacja sygnału na podpasma. Funkcje detekcji poniekąd polegają na redukcji informacji która utrudnia proces znajdowania momentu onset. Możemy użyć funkcji opartych o analizę energii sygnału (RMS i pochodne) jak i analizę widmową. Następnie znajdujemy piki i lokalne minima – onset.
Do poczytania:
http://www.iro.umontreal.ca/~pift6080/H09/documents/presentations/xavier_bello_tutorial.pdf
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.332.989&rep=rep1&type=pdf
https://www.eecs.qmul.ac.uk/~simond/pub/2006/mirex-onset.pdf
Case study:
https://musicinformationretrieval.com/novelty_functions.html
https://www.audiolabs-erlangen.de/resources/MIR/FMP/C6/C6S1_NoveltySpectral.html



Obszar wyszukiwania muzyki Music Information Retrieval (MIR) obejmuje oprogramowanie i rozwiązania sprzętowe w tym zakresie. Łańcuch przetwarzania składa się z urządzeń rejestrujących próbki (mikrofon, smartfon), oprogramowanie zajmujące się przetwarzaniem wstępnym i transmisją zapytania na serwer, jak i usługą chmurową zajmującą się znalezieniem poprawnej odpowiedzi na zadane pytanie. Skuteczność systemu zależy od komunikacji między źródłem danych a systemem. Najprościej ujmując, zadanie polega na przetłumaczeniu opisu semantycznego źródeł dźwięku na język, którym operują systemy komputerowe, pozwalając przy tym na wielowymiarową analizę wykonywaną w ściśle określonym czasie



Automatyczne tagowanie plików audio cieszy się największą popularnością w serwisach muzycznych zajmujących się sprzedażą / udostępnianiem nagrań na przykład na potrzeby filmu, sample itp. Odpowiednie dopasowanie algorytmów pozwala na szybsze wyszukiwanie interesujących klipów. Najczęściej bazują na parametrach MPEG-7. Uwzględniając parametry wysokiego i średniego poziomu jako szukane tagi.
Bez systemów zautomatyzowanych, tagi były przypisywane przez użytkowników (np. last.fm). Pierwsze systemy rekomendacji korzystały z danych opisywanych w ramach „społecznego tagowania" (social tagging, Milicevic et al., 2010; Bischoff, Firan, Nejdl, & Paiu, 2010).
W przypadku systemów automatycznego tagowania, warto zwrócić uwagę na prace Tzanetakis & Cook (2002). Pierwsze dziesięciolecie tego wieku opierało systemy klasyfikacyjne o GMM, SVM, czy AdaBoost. Dużo badań dotyczyło również stopnia rozwinięcia wektora cech – jaka jest optymalna liczba parametrów pomagająca systemowi uzyskać zadaną skuteczność. W okresie tym wyróżniano dwa podejścia – analiza tylko na podstawie parametryzacji, lub parametryzacja plus social tagging.
Najczęściej stosowane parametry: FFT, UTI, MFCC, LPC, MPEG-7, MP, SC, BW, CFRs, RS, MSC, MPCC, BIC, Roll-off, Flux, BOF, ENT, STFT, KLIEP, SCR, ZC, Entropy, LSA, SVD, Timbre, CSML, PARAFAC2, LPCC, MFCC-Delta

https://core.ac.uk/reader/10915160

Figure 1.2 z Markus Schedl, Emilia Gómez, Julián Urbano: Music Information Retrieval: Recent Developments and Applications

5

Segmentacja (Audio Segmentation)

- Dzielenie nagrań na fragmenty – na podstawie zadanych warunków (charakterystyk);
- Stosowane również jako etap przetwarzania wstępnego – uwzględnia również jako dzielenie na ramki – jak i wyszukiwanie aktywności dźwięku;

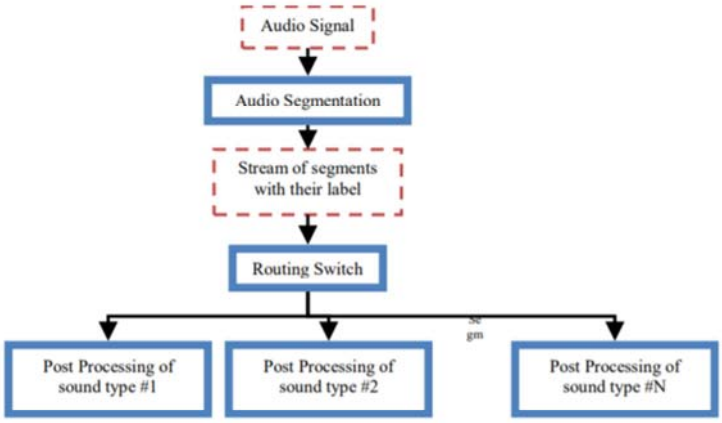


Fig. 1. Automatic audio segmentation and post-processing.

Etap ten pozwala na rozwiniecie możliwości innych systemów, analizując sygnały w mniejszych fragmentach zamiast cały plik na raz. Dzielenie sygnału na jednakowe ramki z zakładkowaniem nie rozwiązuje części problemów.
Stosując segmentacje na podstaie zadanych warunków dokonujemy już wstępnej segregacji próbek dźwięku, podchodząc do systemu rozpoznawania, klasyfikacji jako system kaskadowy (złota zasada, szereg klasyfikatorów da lepszy efekt niż jeden rozbudowany)

http://www.mecs-press.org/ijitcs/ijitcs-v6-n11/IJITCS-V6-N11-1.pdf
Case study: https://www.analyticsvidhya.com/blog/2017/11/heart-sound-segmentation-deep-learning/ (tu przy segmentacji bicia serca)

Przetwadzanie t=f najczęściej prowadzi do uzyskania spektrogramu – po separacji (wycięciu) obraz wraca do postaci czasowej.

Ludzie mają wrodzony filtr separacji źródeł – cocktail-party effect.

http://ijcert.org/ems/ijcert_papers/V3I1103.pdf
Case study: https://github.com/IoSR-Surrey/untwist

7

Śledzenie rytmu (Beat Tracking)

- Śledzenie rytmu
  - Bazować może na *onset detection*
- Nie każdy system działa w czasie rzeczywistym,
- Zastosowania:
  - Automatyczny montaż wideo do podkładu (układani albumów zdjęciowych)
  - Edycja audio
  - Interfejsy człowiek-maszyna

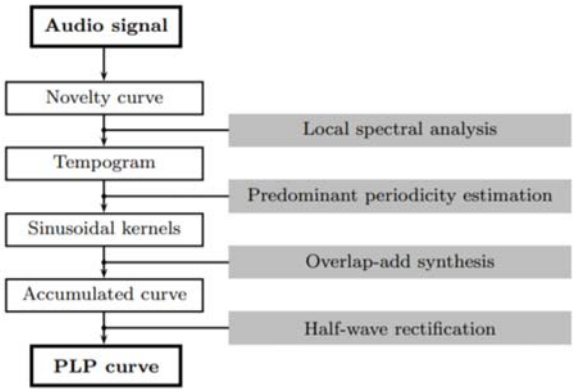


Figure 2.1: Flowchart of the steps involved in the PLP computation.

Do poczytania o temacie https://www.audiolabs-erlangen.de/content/05-fau/professor/00-mueller/01-students/2012_GroschePeter_MusicSignalProcessing_PhD-Thesis.pdf

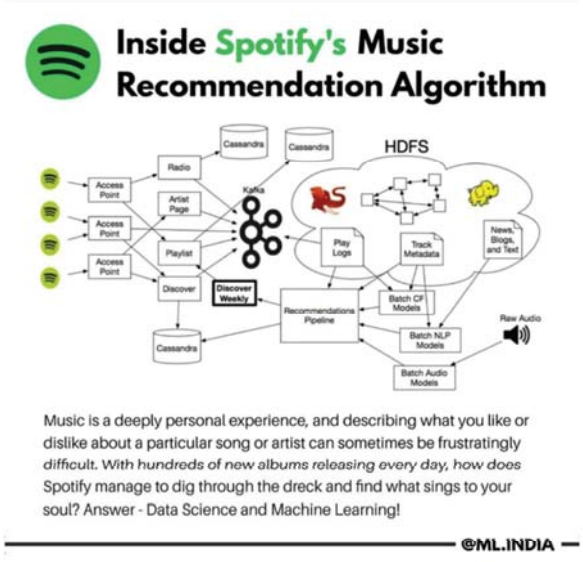
Czas działania systemu powiązany jest z jego dokładnością.

Przykład zastosowania: https://github.com/adamstark/BTrack

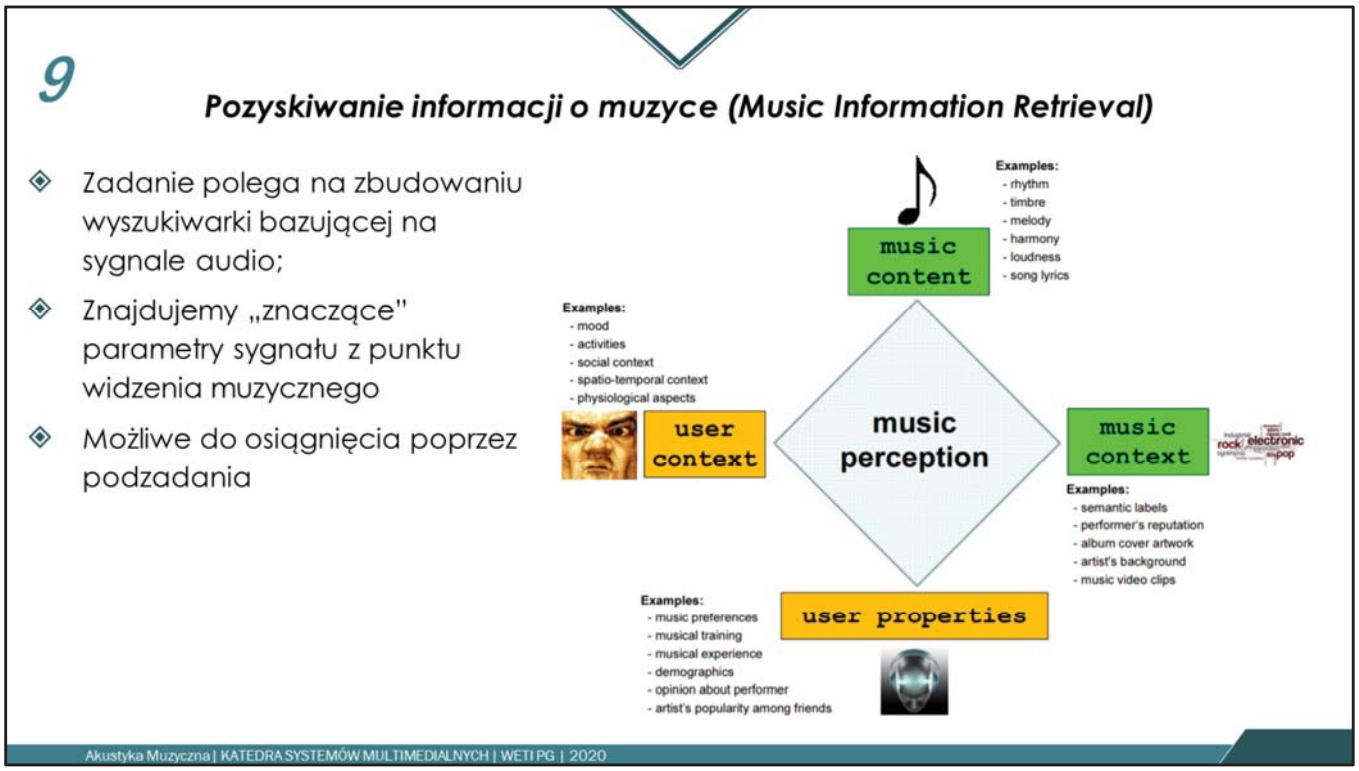
8

Systemy rekomendacji (Music Recomendation)

- Zadanie wspomagające poszerzanie bazy utworów muzycznych spasowanych z naszym gustem.
- Łączy się z już omówionymi zadaniami – automatyczne tagowanie, klasyfikacja)
- Obecnie najbardziej rozwijane w serwisach streamingowych (spotify, saavn, youtube, dawniej last.fm, pandora)



Pierwsze systemy rekomendacji opierały się o proste i szybkie algorytmy (np. regresji). Obecnie w miarę rozwijania baz danych przeniesiono do na deep learning (spotify: https://benanne.github.io/2014/08/05/spotify-cnns.html).
Dane które mogą być stosowane w systemach rekomendacji są dość obszerne. Last.fm bazował na naszych listach, i listach ludzki z którymi mieliśmy wspólne utwory, lub fakt posiadania się w znajomych.
Przegląd stanu z 2017 roku
https://pdfs.semanticscholar.org/7442/c1ebd6c9ceafa8979f683c5b1584d659b728.pdf



Pod działanie tego zadania można również podczepić systemy rekomendacji, czy też n*p.* *Fingerprinting,*

Wymaga takich zadań jak: Analiza tonalna (melodia, harmonia),Analiza rytmu i tempa. By następnie, na podstawie zbudowanej bazy – informacji o danym obiekcie znalezienie innego podobnego obiektu.

http://84.89.139.82/system/files/publications/article_mir_online_0.pdf

10

Transkrypcja (Music Transcription)

- Przetworzenie pliku dźwiękowego na zapis muzyczny (nuty, kod MIDI);
- Wyróżniamy modele automatyczne lub pół-automatyczne wymagające zaangażowania użytkownika
- Stopień zaawansowania zależy dodatkowo od rodzaju sygnału (mono/poli, rodzaj źródła);
- Duże zainteresowanie na rynku

Podobnie jak w przypadku rozpoznawania mowy, zadanie polega na „przepisaniu” utworu na notacje muzyczną.
Łączy w sobie sporo wspomnianych zagadnień (onset, parametryzacja, beat tracking, separacja).
Serwisy z transkrypcją ułatwiają życie mniej muzycznym użytkownikom (nie wymaga od muzyków doskonałego słuchu muzycznego, przyśpiesza proces zapisu utworów zaimprowizowanych, lub tych którymi chcą się zainspirować, nauczyć)

https://ieeexplore.ieee.org/document/1495485

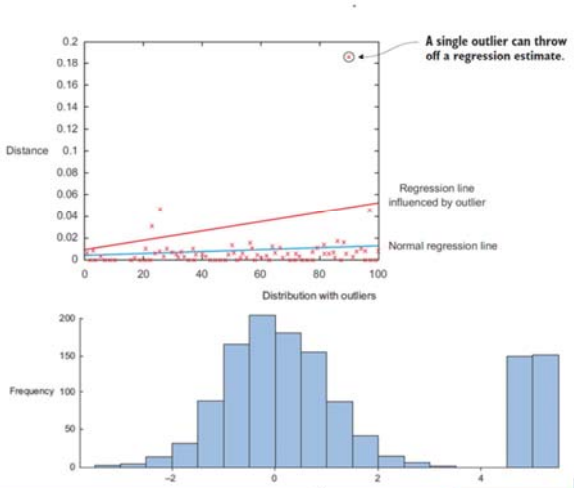
Przykład przeprowadzony w miarę krok po kroku https://youtu.be/9boJ-Ai6QFM

Przykładowe serwisy zapewniające taką usługę:
https://melodyscanner.com/
https://scorecloud.com/
https://www.lunaverus.com/
https://www.mysheetmusictranscriptions.com/


Remember!
Kilka uwag

Data preparation

- Różne parametry uzyskują różne skale wartości (transformacja danych). Należy mieć na uwadze docelowy algorytm klasyfikacji.
- Czyszczenie danych
- Zwiększanie bazy
- Redukcja parametrów
  - Korelacja,
  - PCA
  - LDA, FLDA



Część algorytmów wymaga równych zakresów liczbowych by nie wpłynąć (w pewniej sposób wagowo) na decyzje bez rzeczywistych przesłanek. Część algorytmów analizuje wariancje w grupie – nie należy więc zatasować metod normalizacji wyrównujących wariancje dla każdego parametru.
Czyszczenie danych – usuwamy parametr który brakował w dużej liczbę obiektów, czy wywalamy obiekty które mają braki na parametrach. Nie tylko braki mogą zakłócić nasz pomiar, ale i wartości które znacząco odbiegają od reszty danych. Można wynikać błędnego tagowania, lub zaszumienia próbki. Detekcja „wyrzutków" możliwa jest przez analizę rozkładu wartości
Zwiększanie bazy może zachodzić na plikach – zaszumianie, lub na wartościach parametrów uwzględniając rozkład wartości parametru.
Na pierwszym etapie często tworzymy wszystkie możliwe parametry i ich kombinacje na jakie pozwala nam baza i bibloteki. Nie zawsze należy wykorzystać je wszystkie w procesie uczenia.
Klasyczne metody opierają się o analizę korelacji.