

Reprezentacja wiedzy. Rodzaje  
danych i ich wstępna obróbka.

Metody akwizycji sygnałów  
fonicznych. MPEG7

Zebrała: prof. Bożena Kostek

# Standard MPEG-7

Standard MPEG-Moving Picture and Audio Coding Experts;

MPEG-1 (89r), MPEG-2 (91r), MPEG-4 (95r),  
(MPEG-3 definicja HDTV, MPEG-5, MPEG-6 – nie zdefiniowane)

**MPEG-7** (listopad 2000 r.)

standard MPEG-7, „Multimedia Content Description Interface” - ISO 15938:

- stworzenie języka opisu (ang. Description Definition Language) **zawartości** obrazu, dźwięku, baz multimedialnych
- oraz informacji związanych (**metadata**)
- opis sygnału

# Parametryzacja dźwięków instrumentów muzycznych

---

## Cel

- wyznaczenie wektora cech opisujących dźwięk

## Problemy

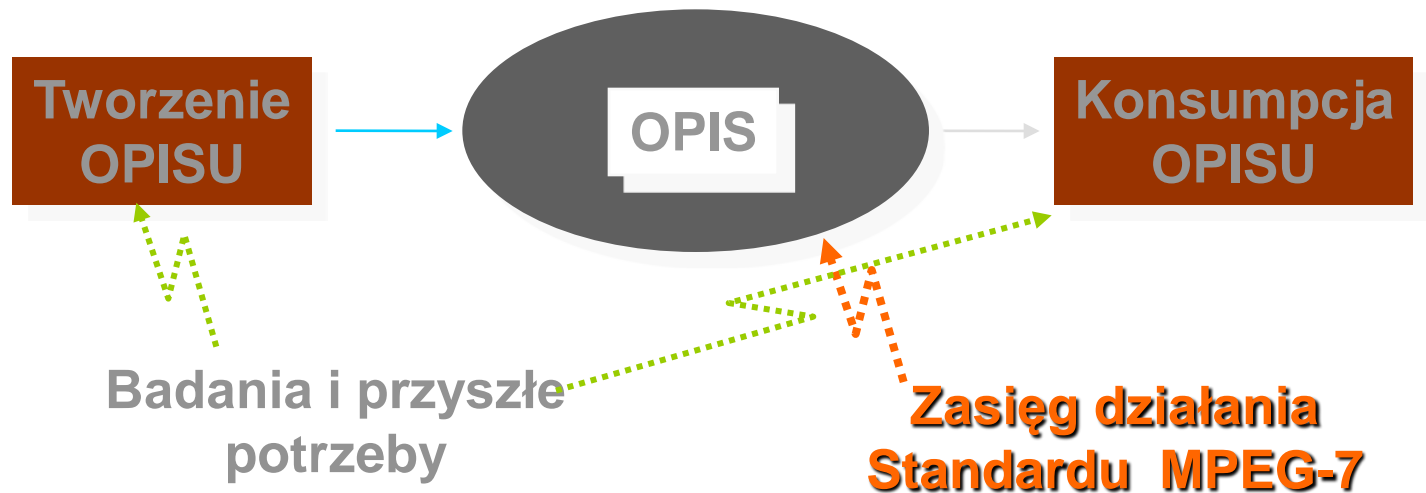
- definicja barwy dźwięku oparta na wrażeniach subiektywnych
- zmienność barwy dźwięku w czasie

# MPEG-7 *Multimedia Content Description Language*

## ISO/IEC 15938: MPEG7

Standard dostarcza technologii do opisu źródeł, które pozwalają na opis zawartości multimedialnej w środowisku multimedialnym

Standard międzynarodowy zatwierdzony we wrześniu 2001R



# **MPEG-7** *Multimedia Content Description Language*

Główne cele:

- opis zawartości multimedialnej
- elastyczność w zarządzaniu danymi
- globalizacja i wewnętrzna kompatybilność zasobów danych

## **MPEG-1 (1992)**

. Pierwszy w historii standard kodowania audio, Warstwy 1-3 (DAB, Worldspace, DVB Internet Audio/"MP3")

## **MPEG-2 (1994)**

. Rozbudowane kodery MPEG-1, zwrócone w kierunku niższych współczynników próbkowania wielokanałowego

## **MPEG-2 AAC (1997)**

. Silniejszy sygnał monofoniczny, kodowanie wielokanałowe

## **MPEG-4 (1999 +)**

. Nowe funkcjonalności (skalowalność, reprezentacja zorientowana obiektowo, interaktywność)

## **MPEG-7 (2001)**

. Standard metadanych (nie kodowania!)

# **MPEG-7** *Multimedia Content Description Language*

## CZĘŚCI

<b>PART I</b>	<b>SYSTEM</b>
<b>PART II</b>	<b>DESCRIPTION DEFINITION LANGUAGE (DDL)</b>
<b>PART III</b>	<b>VISUAL</b>
<b>PART IV</b>	<b>AUDIO</b>
<b>PART V</b>	<b>MULTIMEDIA DESCRIPTION SCHEMES (MDS)</b>
<b>PART VI</b>	<b>REFERENCE SOFTWARE</b>
<b>PART VII</b>	<b>CONFORMANCE TESTING</b>
<b>PART VIII</b>	<b>EXTRACTION AND USE OF MPEG7 DESCRIPTIONS</b>
<b>PART IX</b>	<b>PROFILES</b>
<b>PART X</b>	<b>SCHEMA DEFINITION</b>

# **MPEG-7** GŁÓWNE ELEMENTY STANDARDU

**deskrytory (D, ang. *Descriptors*)** – sposób opisu poszczególnych cech (elementów metadanych),

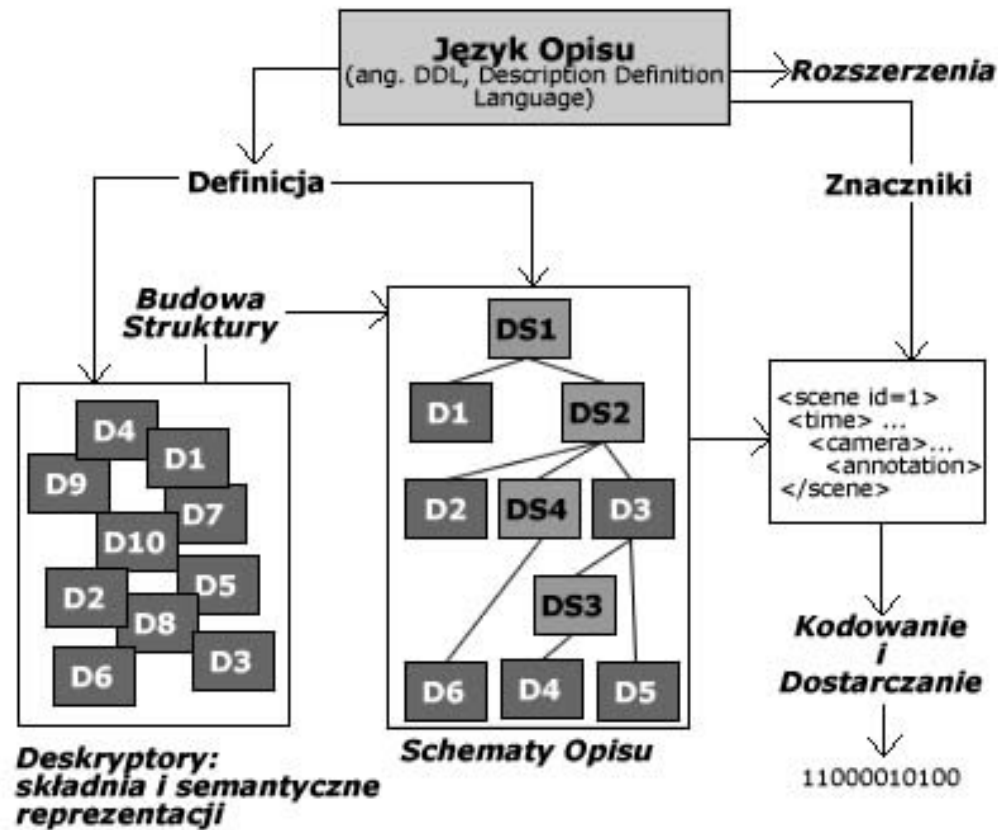
**schematy opisu (DS, ang. *Description Schemes*)** – sposób opisu relacji (struktury i składni) między deskryptorami (również pomiędzy różnymi schematami deskryptorów),

**język definicji deskryptorów (DDL, ang. *Description Definition Language*)** – język do tworzenia opisów (również do tworzenia nowych schematów lub deskryptorów),

**schematy klasyfikacji (CS, ang. *classification schema*)** – pojęcia i znaczenia używane do opisu danych

**narzędzia systemowe (*system tools*)** – przechowywanie i transmisja danych,

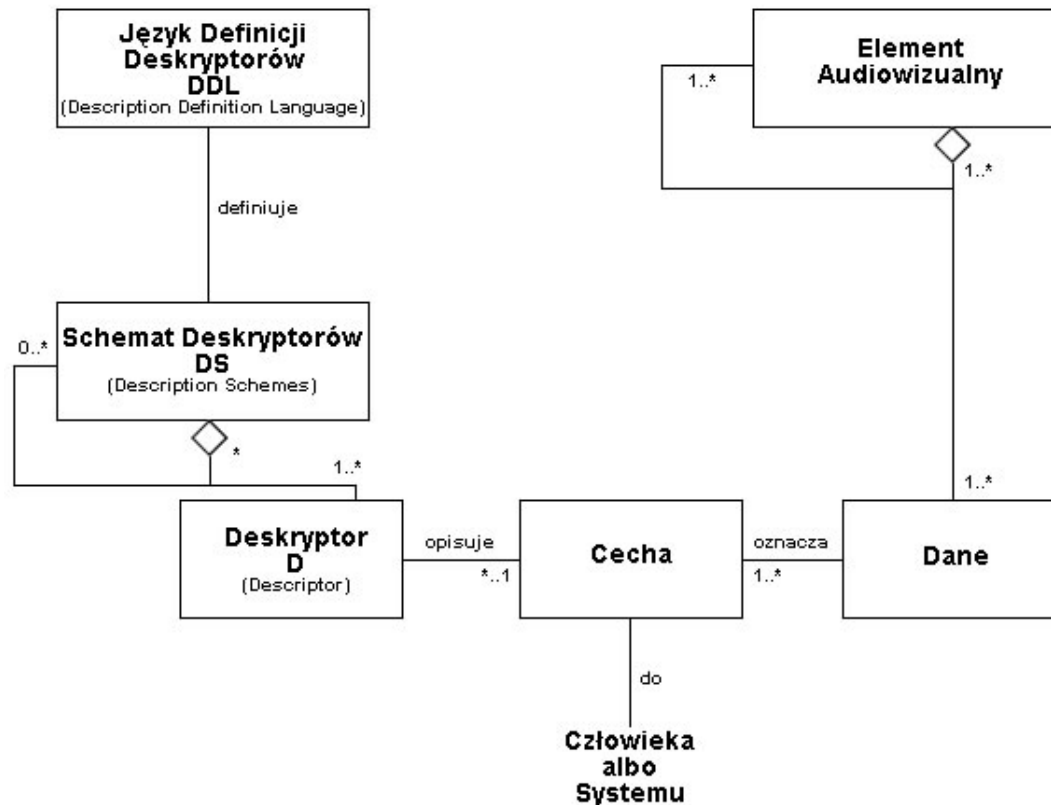
## MPEG-7 GŁÓWNE ELEMENTY STANDARDU





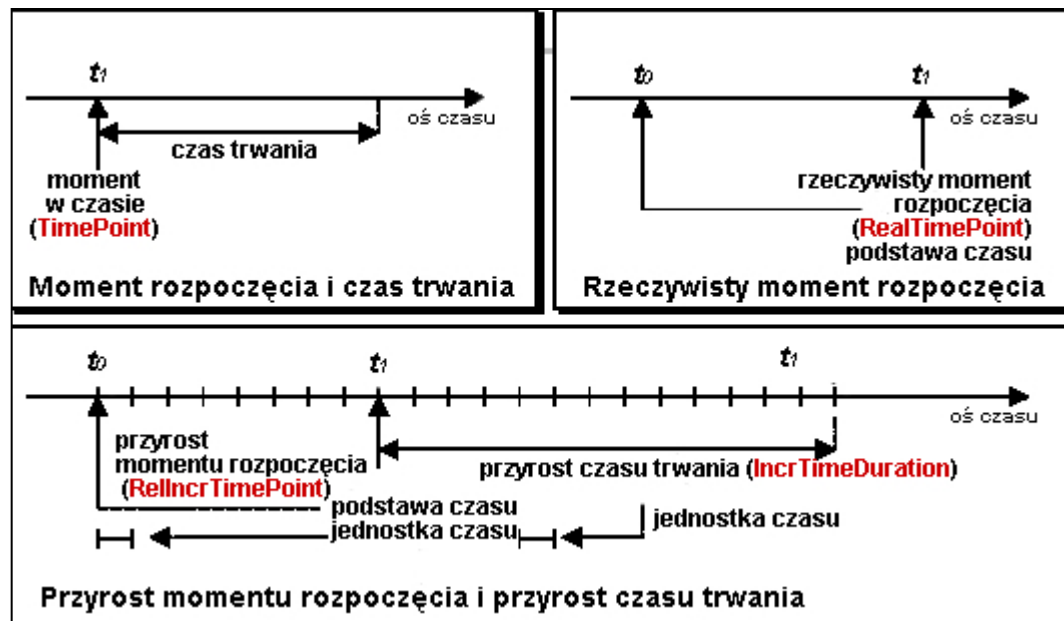
# MPEG-7 ZWIĄZKI POMIĘDZY Ds I DSs

Reprezentacja UML (ang. *Unified Modeling Language*) możliwych związków pomiędzy *Ds* i *DSs*



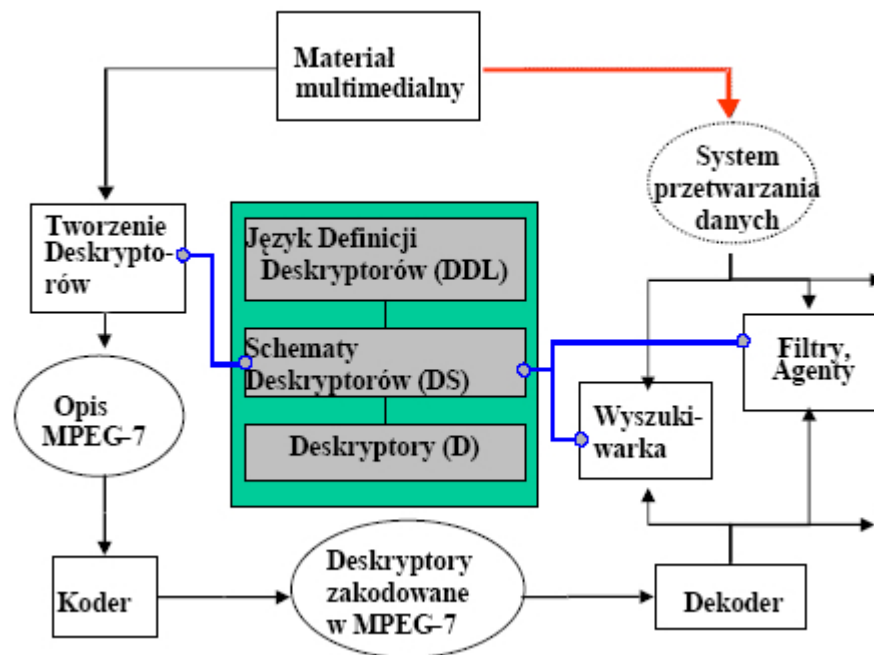
# MPEG-7 CZASOWE SCHEMATY DESKRYPTORÓW DS

Przegląd czasowych schematów opisu deskryptorów (DSs)



## MPEG-7 PRZYKŁADOWE APLIKACJE

Reprezentacja możliwych aplikacji przy użyciu MPEG-7



Lewa strona diagramu przedstawia sposoby przypisywania danych, a prawa – sposoby ich pozyskiwania.

W kwadratach zaprezentowano narzędzia do przetwarzania (kodowania i dekodowania), a w elipsach – elementy statyczne, takie jak OPIS.

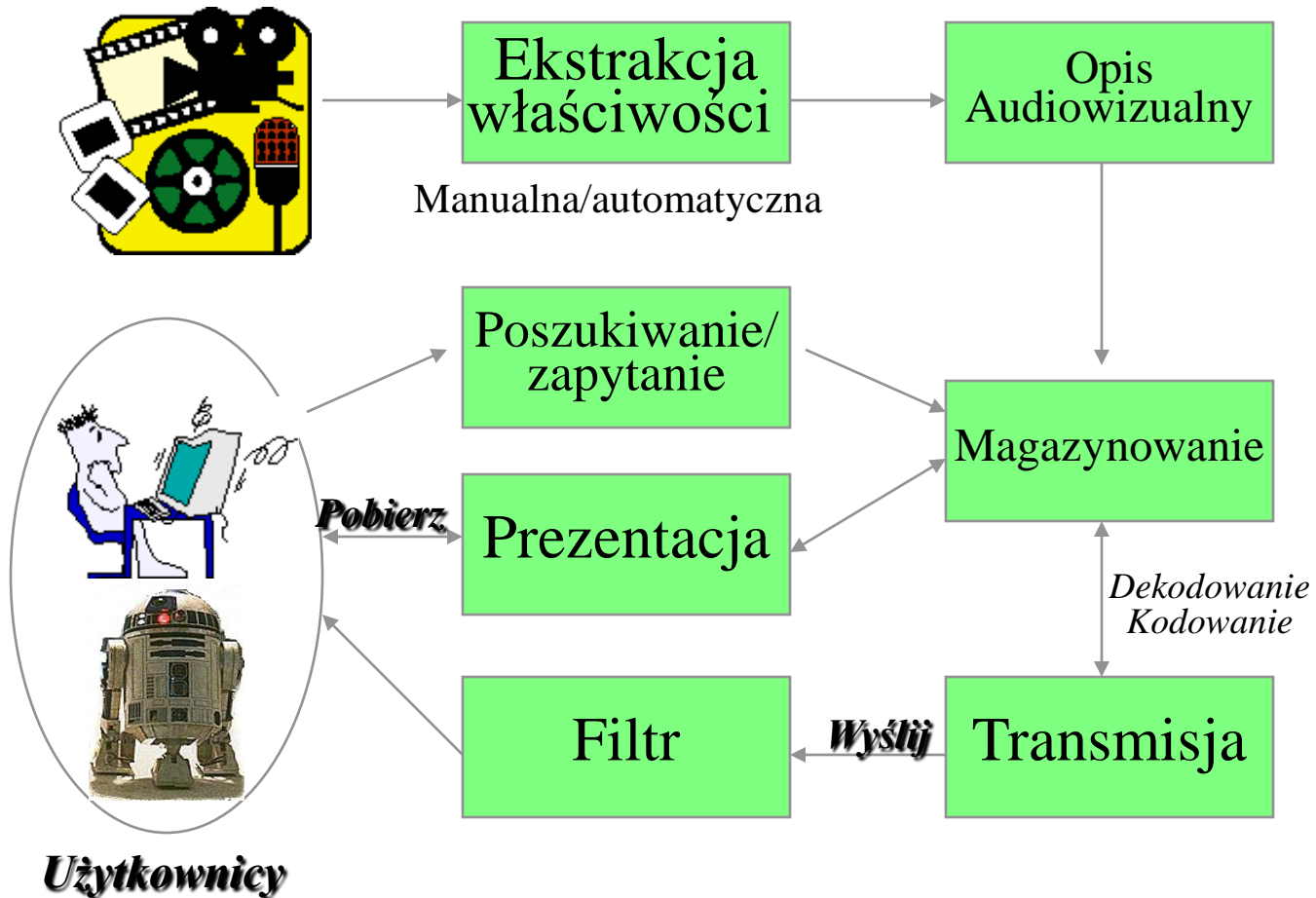
Przykładowo kwadrat z informacją: „Tworzenie Deskryptorów”, to opis mechanizmu tworzenia produkcji „Opisu MPEG-7”, która przedstawiona jest w elipsie.

# **MPEG-7** ZASTOSOWANIA

- Przechowywanie i przeszukiwanie baz danych audiowizualnych (obrazy, film, archiwa radiowe)
- Selekcja mediów transmisyjnych (radio, programy TV)
- Kontrola (kontrola ruchu, łańcuchy produkcyjne)
- E-handel and Tele-zakupy (szukanie odzieży)
- Odczyt zdalny (kartografia, ekologia, zarządzanie zasobami naturalnymi)
- Medycyna – obróbka sygnałów fonicznych
- Kultura (muzea, galerie sztuki); Rozrywka (wyszukiwanie gier, plików muzycznych)
- Dziennikarstwo (wyszukiwanie zdarzeń, osób)
- Personalized news service on Internet (push media filtering)
- Inteligentna prezentacje multimedialne
- Aplikacje edukacyjne

## MPEG-7 ZASTOSOWANIA

Drogi przepływu informacji



# MPEG-7

## **PART I** Systems

binarny format kodowania deskryptorów MPEG-7  
i architektury terminala

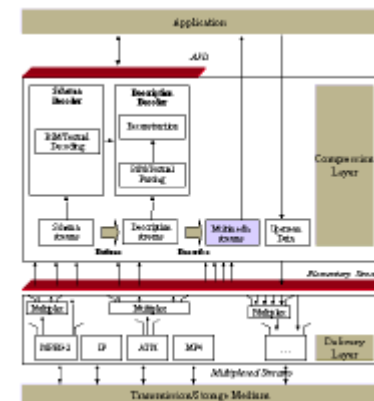
# MPEG-7 SYSTEMS

Pierwsza część standardu charakteryzuje:

- architekturę terminala
- interfejsy normatywne

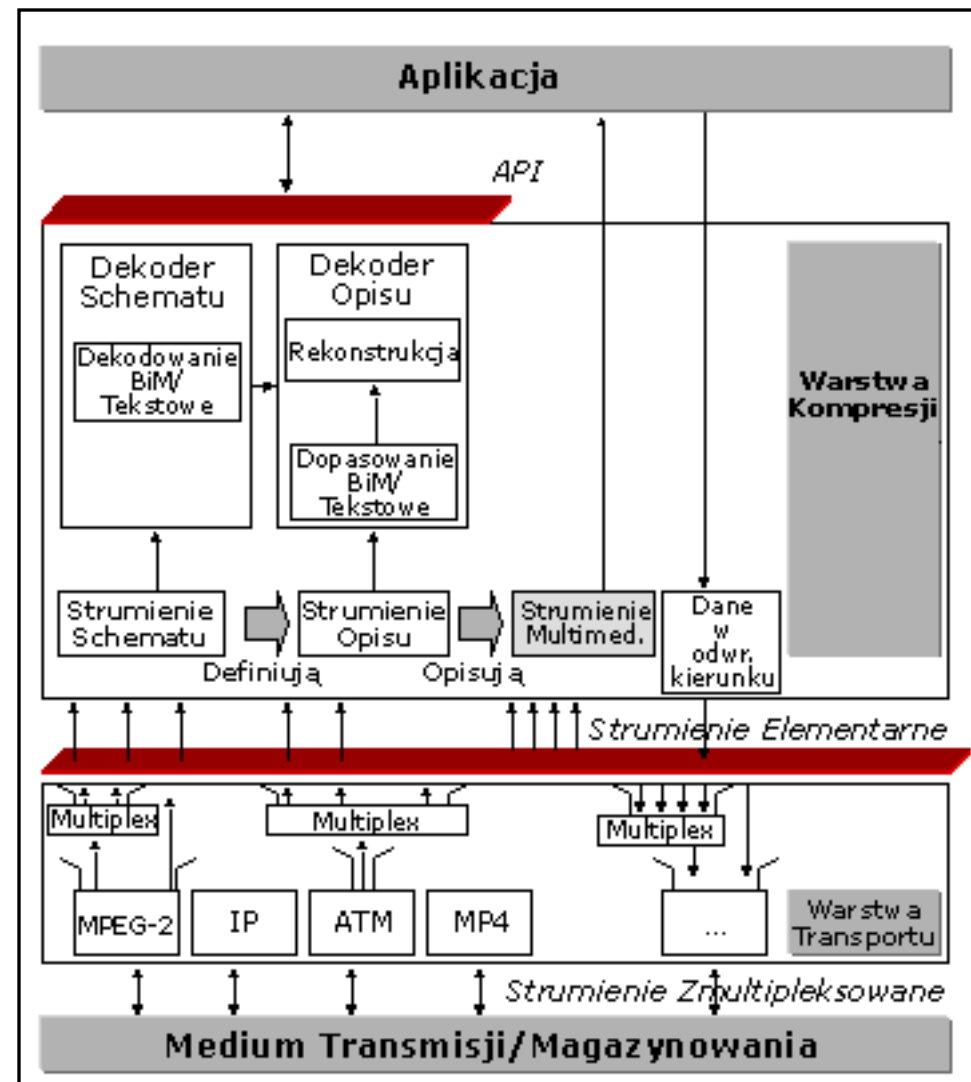
*Terminal* – jednostka wykorzystująca zakodowaną reprezentację zawartości multimedialnej; może odnosić się do wolnostojącej aplikacji lub być częścią systemu aplikacji:

- aplikacja
- warstwa kompresji
- warstwa transportu
- medium transmisyjne/magazynujące



## MPEG-7 SYSTEMS

### STRUKTURA TERMINALA

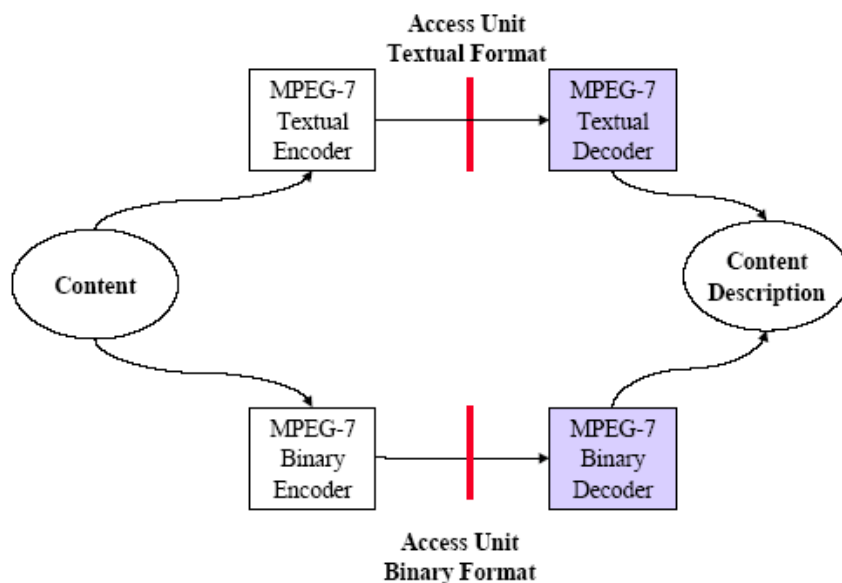




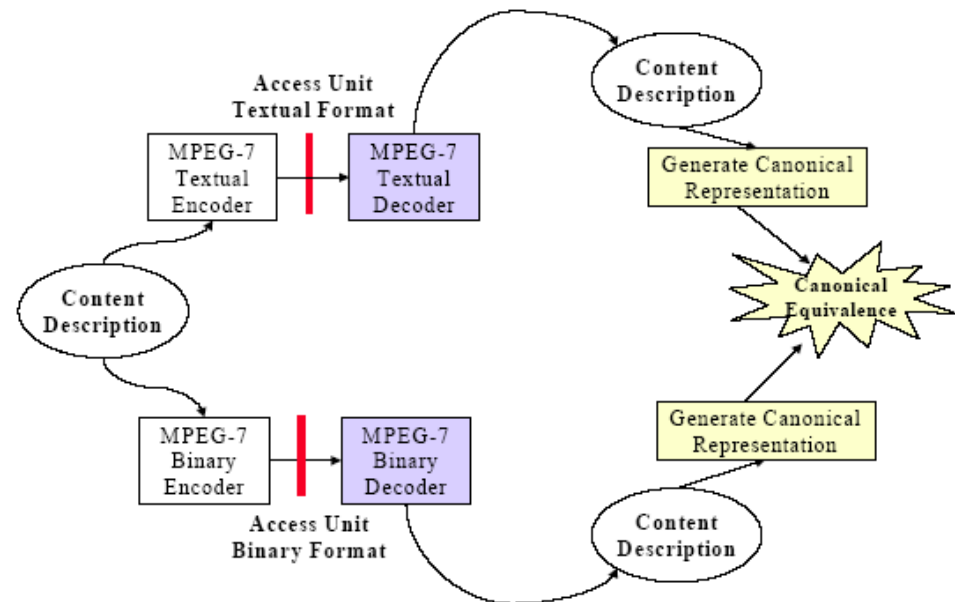
## MPEG-7 SYSTEMS

Dwa rodzaje interfejsów normatywnych:

- tekstowy
- binarny



Ocena poprawności systemu



# MPEG-7

## **PART II** Description Definition Language (DDL)

język definiujący składnię narzędzi opisu w MPEG-7  
i nowe schematy opisu

# MPEG-7 DDL

język DDL służy do tworzenia deskryptorów i ich schematów, a więc do opisu metadanych.

Wg standardu MPEG-7, DDL wykorzystuje język **XML**.

Wykorzystywany jest też standard **XML-Schema**:

- struktura dokumentu MPEG-7,
- dane składowe i ich typy,
- rozszerzenia standardu *XML-Schema* wprowadzone przez MPEG-7.

Obok: nagłówek dokumentu zgodnego ze standardem MPEG7

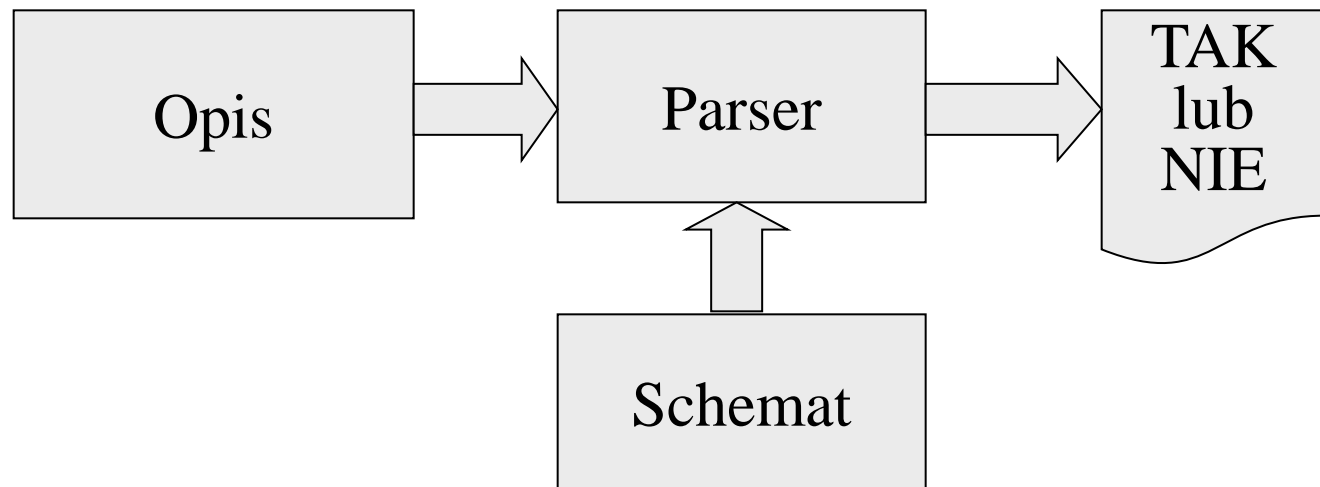
```
<?xml version="1.0" encoding="iso-8859-1?">
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
  xmlns:xml="http://www.w3.org/XML/1998/namespace"
  xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001
Mpeg7-2001.xsd">

<!-- Tu wstaw zawartość MPEG-7 -->

</Mpeg7>
```

# MPEG-7 DDL - Parser

Parser DDL – oprogramowanie służące do weryfikacji  
poprawności opisu  
(porównanie dokumentu ze schematem)



# MPEG-7 DDL

Projekt DDL charakteryzuje podstawę prac nad standardem MPEG-7. Stanowi silną podporę opisową niezbędną dla użytkowników tworzących swoje własne schematy opisu (DS) i deskrytory (D). W tym celu DDL musi spełniać poniższe wymagania:

. **Możliwości komponowania** - daje możliwość tworzenia własnych DSs i Ds, pamiętając o możliwości budowania DS z różnorodnych Ds.

. **Możliwości transformacji** – DDL pozwala na ponowne użycie, rozszerzanie i dziedziczenie istniejących Ds i DSs.

. **Unikatowa identyfikacja** – dostarcza mechanizm pozwalający na jednoznaczną identyfikację DSs i Ds

. **Typy danych** – dostarcza zestawu podstawowych typów danych, takich jak tekst, zmienna, dane rzeczywiste, czas/znacznik czasu, wersja, aby właściwie opisać typy danych, składające się na D czy DS.

. **Relacje w obrębie DS i pomiędzy DSs** – DDL dostarcza możliwość wyrażania relacji pomiędzy DSs i pomiędzy elementami jednego DS.; DDL wyraża semantyczne dane dotyczące tych relacji, np. relacje przestrzenne, czasowe, strukturalne, pojęciowe.

. **Relacje pomiędzy D i danymi** – DDL dostarcza rozbudowany model do tworzenia odniesień w obrębie jednego lub pomiędzy kilkoma Ds i opisanymi danymi.

# MPEG-7

## **PART III** Audio

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Podstawowe (Basic)**

*AudioWaveform, AudioPower*

- **Podstawowe deskrytory widmowe (Basic Spectral)**

*AudioSpectrumEnvelope, AudioSpectrumCentroid, AudioSpectrumSpread, AudioSpectrumFlatness*

- **Parametry Sygnału (Signal Parameters)**

*AudioFundamentalFrequency, AudioHarmonicity*

- **Parametry czasowe barwy dźwięku (Timbral Temporal)**

*LogAttackTime, TemporalCentroid*

- **Parametry widmowe barwy dźwięku (Timbral Spectral)**

*SpectralCentroid, HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, HarmonicSpectralVariation*

- **Deskrytory dynamicznego widma sygnału (Spectral Basis)**

*AudioSpectrumBasis, AudioSpectrumProjection*

- **Cisza (Silence)**

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Podstawowe (Basic)**

*AudioWaveform, AudioPower*

Dwa podstawowe deskryptory audio są próbkowanymi w czasie wartościami skalarnymi stosowanymi w celu wizualizacji przebiegu sygnału.

. *AudioWaveform* – określa zakres obwiedni przebiegu czasowego sygnału (jego minimum i maksimum)

. *AudioPower* – uśredniona w czasie moc sygnału



# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

### ▪ Podstawowe deskryptory widmowe (Basic Spectral)

*AudioSpectrumEnvelope, AudioSpectrumCentroid, AudioSpectrumSpread, AudioSpectrumFlatness*

#### . *AudioSpectrumEnvelope* (ASE):

- opisuje widmo sygnału audio w funkcji częstotliwości przedstawionej w skali logarytmicznej
- graficzna reprezentacja tego parametru to spektrogram
- podstawowe narzędzie w procesie rozpoznawania dźwięków, wyszukiwania i porównań
- domyślny zakres częstotliwości dla ASE: od 62.5 Hz (low-edge) do 16 kHz (high-edge)
- obie granice zakresu ASE muszą spełniać zależność:

$$edge = 2^{rm} \times 1\text{KHz}$$

gdzie:  $r$  – rozdzielczość w oktawach,  $m$  należy do  $\mathbf{Z}$

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Podstawowe deskryptory widmowe (Basic Spectral)**

*AudioSpectrumEnvelope, AudioSpectrumCentroid, AudioSpectrumSpread, AudioSpectrumFlatness*

- . *AudioSpectrumCentroid (ASC)*

- określa środek ciężkości widma mocy w skali logarytmicznej, wyliczany na podstawie wzoru:

$$C = \frac{\sum_n \log_2(f(n)/1000)P'_x(n)}{\sum_n P'_x(n)}$$

gdzie oznaczony jest jako moc  $P'_x(n)$  powiązana z częstotliwością  $f(n)$

- opisuje widmo sygnału audio w funkcji częstotliwości przedstawionej w skali logarytmicznej

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

### ▪ Podstawowe deskryptory widmowe (Basic Spectral)

*AudioSpectrumEnvelope, AudioSpectrumCentroid, AudioSpectrumSpread, AudioSpectrumFlatness*

#### . *AudioSpectrumSpread* (ASS)

- opisuje drugi moment statystyczny widma mocy
- jest efektywnym deskryptorem kształtu widma mocy
- informuje czy moc jest skupiona w otoczeniu częstotliwości swojego środka masy, czy rozrzucona w całym zakresie widma
- deskryptor wyliczany jest na podstawie wzoru:

$$S = \sqrt{\frac{\sum_n (\log_2(f(n)/1000) - C)^2 P'_x(n)}{\sum_n P'_x(n)}}$$

gdzie  $C$ ,  $P'_x(n)$  i  $f(n)$  to wartości pochodzące ze wzoru *AudioSpectrumCentroid Descriptor*

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Podstawowe deskryptory widmowe (Basic Spectral)**

*AudioSpectrumEnvelope, AudioSpectrumCentroid, AudioSpectrumSpread, AudioSpectrumFlatness*

. *AudioSpectrumFlatness* (ASF)

- niesie informację o dewiacji charakterystyki widmowej od płaskiego kształtu w zadanych pasmach częstotliwości

- opisany jest stosunkiem współczynników widma średniej geometrycznej i arytmetycznej:

$$SFM_b = \frac{\sqrt{\prod_{i=il(b)}^{ih(b)} c(i)}}{\frac{1}{ih(b) - il(b) + 1} \sum_{i=il(b)}^{ih(b)} c(i)}$$

gdzie  $ih$ ,  $il$  - najwyższy i najniższy współczynnik widma mocy,  $c(i)$  - pasmo

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Parametry sygnału (Signal Parameters)**

*AudioFundamentalFrequency, AudioHarmonicity*

Dwa deskryptory parametrów sygnału zwykle stosowane są w sygnałach okresowych lub kwazi-okresowych

. *AudioFundamentalFrequency* za pomocą algorytmu śledzenia wierzchołka (ang. *pith tracking*) określa podstawową częstotliwość audio

. *AudioHarmonicity* pozwala określić stopień harmonicznego widma

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Parametry czasowe barwy dźwięku (Timbral Temporal)**

*LogAttackTime, TemporalCentroid*

Oba deskryptory są niezwykle przydatne do opisu barwy dźwięku w muzyce

- . *LogAttackTime*

- charakteryzuje czas narastania obwiedni sygnału (od ciszy do momentu osiągnięcia maksimum) zdefiniowanym w skali logarytmicznej

- na jego podstawie odróżnić można od siebie nagłe i delikatnie rozpoczynające się dźwięki:

$$LAT = \log_{10}(T_1 - T_0)$$

gdzie  $T_0$  to początek, a  $T_1$  – koniec transjentu wejściowego

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Parametry czasowe barwy dźwięku (Timbral Temporal)**

*LogAttackTime, TemporalCentroid*

Oba deskryptory są niezwykle przydatne do opisu barwy dźwięku w muzyce

. *TemporalCentroid* (TC) – wyznacza środek ciężkości obwiedni sygnału – punkt skupienia energii sygnału:

$$TC = \frac{\sum_{n=1}^{length(SEnv)} n / sr * SEnv(n)}{\sum_{n=1}^{length(SEnv)} SEnv(n)}$$

gdzie:

$SEnv(n)$  – obwiednia sygnału

$sr$  – częstotliwość próbkowania (ang. *sampling rate*)

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Parametry widmowe barwy dźwięku (Timbral Spectral)**

*SpectralCentroid, HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, HarmonicSpectralVariation*

. *SpectralCentroid* (SC) – oznacza środek ciężkości widma mocy wyrażony w Hz. Parametr ten jest skorelowany z subiektywnym wrażeniem ostrości dźwięku (ang. *sharpness*)



# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Parametry widmowe barwy dźwięku (Timbral Spectral)**

*SpectralCentroid, HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, HarmonicSpectralVariation*

. *HarmonicSpectralCentroid* (HSC)

- wyliczany jest jako średnia występowania w segmencie dźwięku chwilowego środka ciężkości widma (IHSC):

$$HSC = \frac{\sum_{frame=1}^{nb\_frames} IHSC(frame)}{nb\_frames}$$

gdzie *nb\_frames* - liczba ramek w segmencie dźwięku

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Parametry widmowe barwy dźwięku (Timbral Spectral)**

*SpectralCentroid, HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, HarmonicSpectralVariation*

- . *HarmonicSpectralCentroid* (HSC)

- chwilowy środek ciężkości widma harmonicznego (IHSC) wyliczany jest jako ważona wartość średnia częstotliwości harmonicznych prążków widma, ważona według ich amplitudy:

$$IHSC(frame) = \frac{\sum_{harmono=1}^{nb\_harmono} f(frame, harmo)A(frame, harmo)}{\sum_{harmono=1}^{nb\_harmono} A(frame, harmo)}$$

gdzie:

*nb\_harmo* - liczba harmonicznych wierzchołków wziętych pod uwagę w analizie

*f(frame, harmo)* - częstotliwość wierzchołków harmonicznych w kilku ramkach

*A(frame, harmo)* - amplituda wierzchołków harmonicznych w kilku ramkach

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Parametry widmowe barwy dźwięku (Timbral Spectral)**

*SpectralCentroid, HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, HarmonicSpectralVariation*

. *HarmonicSpectralDeviation* (HSD)

- opisuje różnicę między amplitudami prążków harmonicznych a obwiednią widma

- wyliczany jest jako wartość średnia chwilowej dewiacji widma harmonicznego (IHSD) przez cały czas trwania segmentu:

$$HSD = \frac{\sum_{frame=1}^{nb\_frames} IHSD(frame)}{nb\_frames}$$

gdzie *nb\_frames* - liczba ramek w segmencie dźwięku

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Parametry widmowe barwy dźwięku (Timbral Spectral)**

*SpectralCentroid, HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, HarmonicSpectralVariation*

- . *HarmonicSpectralDeviation* (HSD)

- chwilowa dewiacja widma harmonicznego wyliczana jest jako odchylenie logarytmu amplitud prążków od wartości średniej:

$$IHSD(frame) = \frac{\sum_{harmono=1}^{nb\_harmono} |\log_{10}(A(frame, harmo)) - \log_{10}(SE(frame, harmo))|}{\sum_{harmono=1}^{nb\_harmono} \log_{10}(A(frame, harmo))}$$

gdzie:

*nb\_harmo* - liczba harmoniczných wierzchołków wziętych pod uwagę w analizie

*SE(frame, harmo)* - obwiednia widma dźwięku

*A(frame, harmo)* - amplituda wierzchołków harmoniczných w kilku ramkach

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Parametry widmowe barwy dźwięku (Timbral Spectral)**

*SpectralCentroid, HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, HarmonicSpectralVariation*

. *HarmonicSpectralSpread* (HSS)

- opisuje rozkład harmoniczny w widmie sygnału
- zdefiniowany jest jako odchylenie standardowe amplitud harmonicznych znormalizowane przez wartość *HarmonicSpectralCentroid*:

$$HSS = \frac{\sum_{frame=1}^{nb\_frames} IHSS(frame)}{nb\_frames}$$

gdzie *nb\_frames* - liczba ramek w segmencie dźwięku

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Parametry widmowe barwy dźwięku (Timbral Spectral)**

*SpectralCentroid, HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, HarmonicSpectralVariation*

- . *HarmonicSpectralSpread* (HSS)

- Rozkład harmoniczných w widmie harmoniczným wyliczany jest jako średnia chwilowych rozkładów harmoniczných w widmie (IHSS) przez cały czas trwania segmentu:

$$IHSS(frame) = \frac{1}{IHSC(frame)} \sqrt{\frac{\sum_{harmono=1}^{nb\_harmono} A^2(frame, harmo) [f(frame, harmo) - IHSC(frame)]^2}{\sum_{harmono=1}^{nb\_harmono} A^2(frame, harmo)}}$$

gdzie:

$nb\_harmono$  - liczba harmoniczných wierzchołków wziętych pod uwagę w analizie

$f(frame, harmo)$  - częstotliwość wierzchołków harmoniczných w kilku ramkach

$A(frame, harmo)$  - amplituda wierzchołków harmoniczných w kilku ramkach

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Parametry widmowe barwy dźwięku (Timbral Spectral)**

*SpectralCentroid, HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, HarmonicSpectralVariation*

. *HarmonicSpectralVariation* (HSV)

- znormalizowany współczynnik korelacji amplitud harmoniczných z dwóch sąsiednich ramek sygnału:

$$HSV = \frac{\sum_{frame=1}^{nb\_frames} IHSV(frame)}{nb\_frames}$$

gdzie *nb\_frames* - liczba ramek w segmencie dźwięku

# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Parametry widmowe barwy dźwięku (Timbral Spectral)**

*SpectralCentroid, HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, HarmonicSpectralVariation*

. *HarmonicSpectralVariation* (HSV)

- IHSV:

$$IHSV(frame) = 1 - \frac{\sum_{harmo=1}^{nb\_harmo} A(frame-1, harmo)A(frame, harmo)}{\sqrt{\sum_{harmo=1}^{nb\_harmo} A^2(frame-1, harmo)} \sqrt{\sum_{harmo=1}^{nb\_harmo} A^2(frame, harmo)}}$$

gdzie:

*nb\_harmo* - liczba harmoniczných wierzchołków wziętych pod uwagę w analizie

*A(frame, harmo)* - amplituda wierzchołków harmoniczných w kilku ramkach



# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

- **Deskryptory dynamicznego widma sygnału (Spectral Basic)**

*AudioSpectrumBasis, AudioSpectrumProjection*

Oba deskryptory tej grupy niosą informację o dynamice charakterystyki widmowej.

*AudioSpectrumBasis* (funkcje bazowe widma) *AudioSpectrumProjection* (funkcje przekształcające) wyznaczane są na podstawie kilku pierwszych kolumn macierzy **V**:

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

Mogą posłużyć do procesu automatycznego rozpoznawania dźwięków w oparciu o modele probabilistyczne.

# MPEG-7

GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU

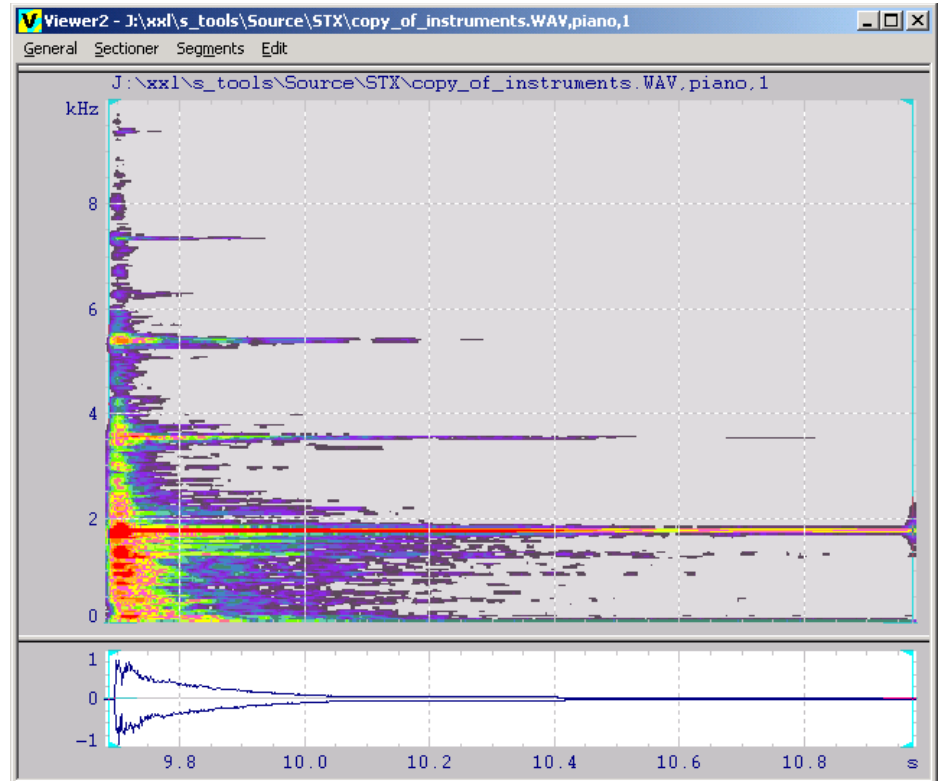
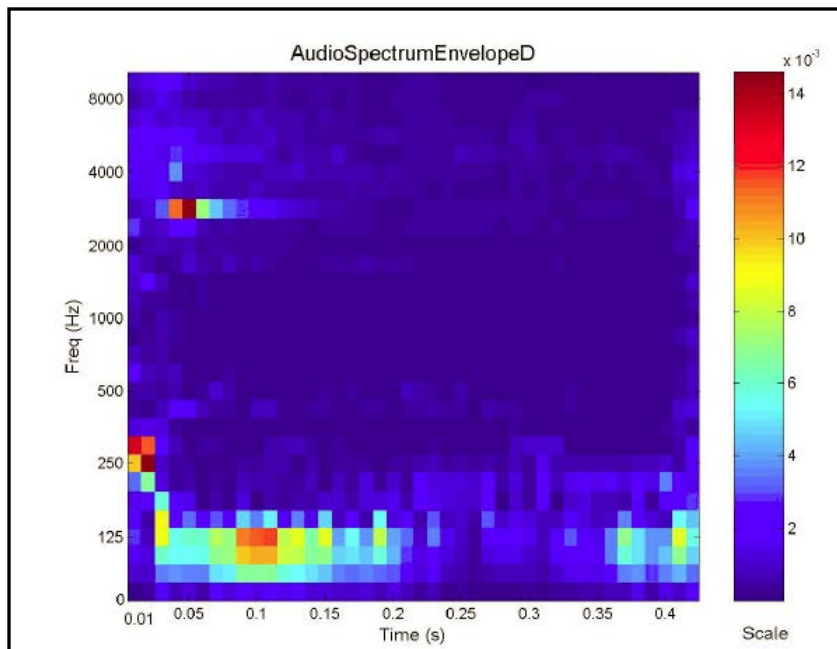
- **Cisza (Silence)**

Deskryptor opisujący fragment nie zawierający żadnych istotnych dźwięków. Jest użyteczny przy segmentacji dźwięku oraz jako parametr ostrzegający przed próbą przetwarzania danego fragmentu.

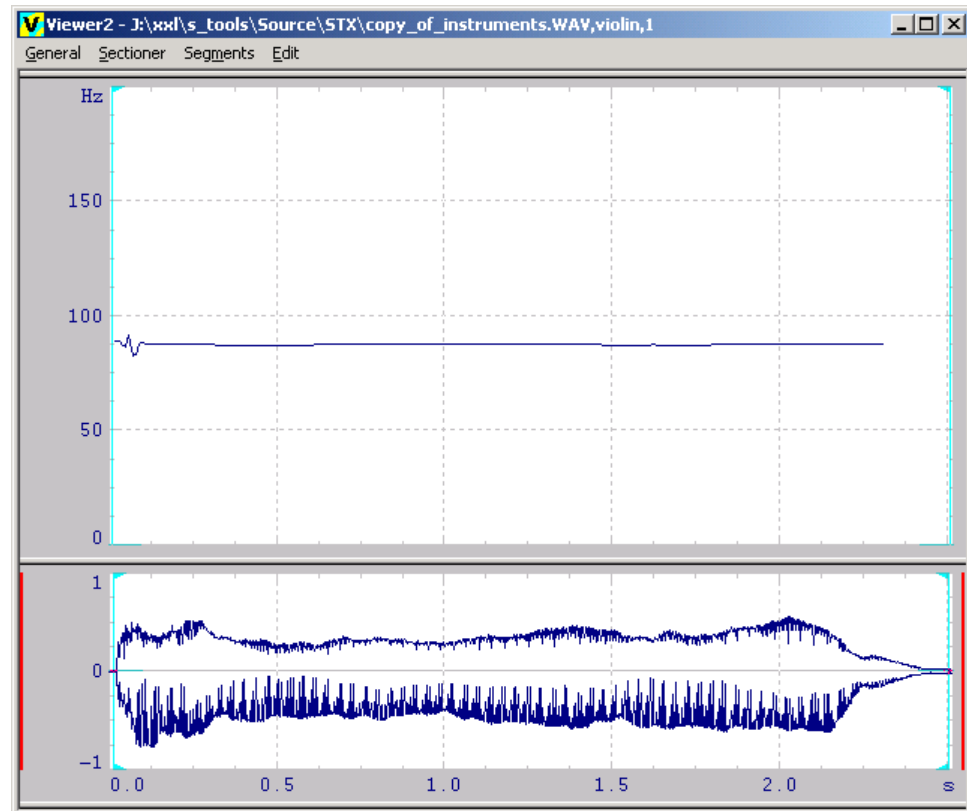
- **Basic (Podstawowe)**  
AudioWaveform, AudioPower



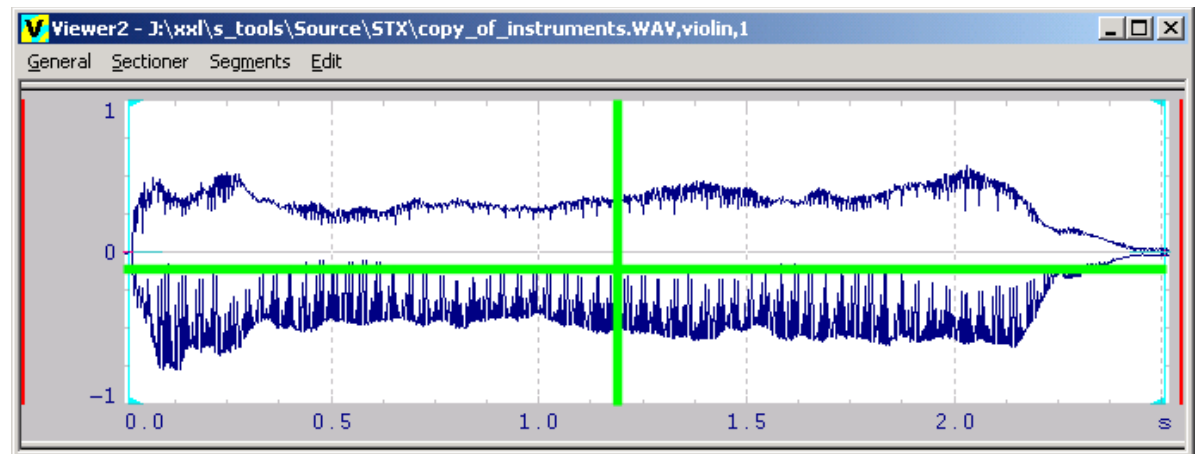
- **Basic (Podstawowe)**  
AudioWaveform, AudioPower
- **Basic Spectral (Podstawowe deskryptory widmowe)**  
AudioSpectrumEnvelope, AudioSpectrumCentroid, AudioSpectrumSpread, AudioSpectrumFlatness



- **Basic (Podstawowe)**  
AudioWaveform, AudioPower
- **Basic Spectral (Podstawowe deskryptory widmowe)**  
AudioSpectrumEnvelope, AudioSpectrumCentroid, AudioSpectrumSpread, AudioSpectrumFlatness
- **Signal Parameters (Parametry Sygnału)**  
AudioHarmonicity, AudioFundamentalFrequency



- **Basic (Podstawowe)**  
AudioWaveform, AudioPower
- **Basic Spectral (Podstawowe deskryptory widmowe)**  
AudioSpectrumEnvelope, AudioSpectrumCentroid, AudioSpectrumSpread, AudioSpectrumFlatness
- **Signal Parameters (Parametry Sygnału)**  
AudioHarmonicity, AudioFundamentalFrequency
- **Timbral Temporal (Parametry czasowe barwy dźwięku)**  
LogAttackTime, TemporalCentroid



- **Basic (Podstawowe)**  
AudioWaveform, AudioPower
- **Basic Spectral (Podstawowe deskryptory widmowe)**  
AudioSpectrumEnvelope, AudioSpectrumCentroid, AudioSpectrumSpread, AudioSpectrumFlatness
- **Signal Parameters (Parametry Sygnału)**  
AudioHarmonicity, AudioFundamentalFrequency
- **Timbral Temporal (Parametry czasowe barwy dźwięku)**  
LogAttackTime, TemporalCentroid
  
- **Timbral Spectral (Parametry widmowe barwy dźwięku)**  
SpectralCentroid, HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, HarmonicSpectralVariation

- **Basic (Podstawowe)**

AudioWaveform, Audio

- **Basic Spectral (P**

AudioSpectrumEnvelope

AudioSpectrumFlatness

- **Signal Parameters**

AudioHarmonicity, Audio

- **Timbral Temporal**

LogAttackTime, Temporal



- **Timbral Spectral (Parametry widmowe barwy dźwięku)**

SpectralCentroid, HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, HarmonicSpectralVariation



- **Basic (Podstawowe)**

AudioWaveform, AudioPower

- **Basic Spectral (Podstawowe deskryptory widmowe)**

AudioSpectrumEnvelope, AudioSpectrumFlatness

- **Signal Parameters (Parametry)**

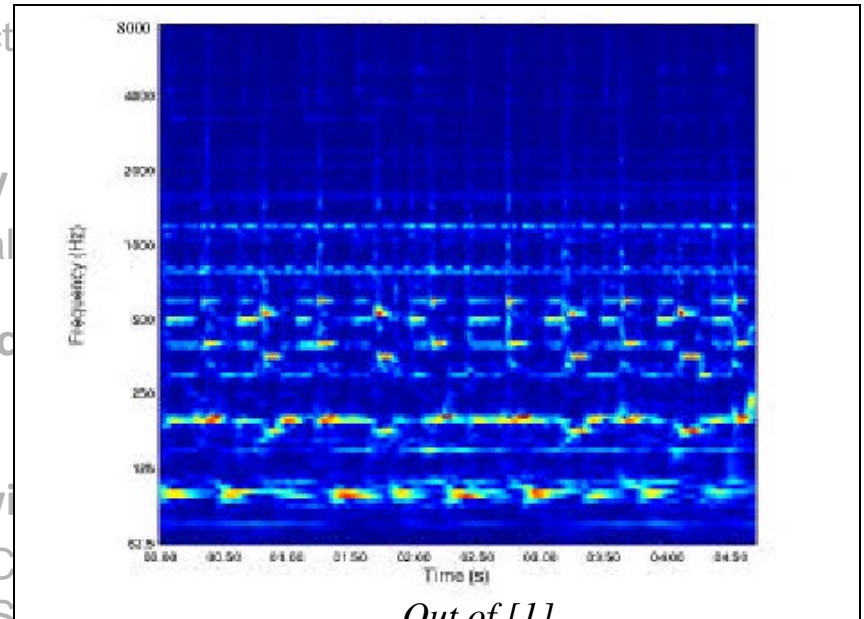
AudioHarmonicity, AudioFundamental

- **Timbral Temporal (Parametry dźwiękowe)**

LogAttackTime, TemporalCentroid

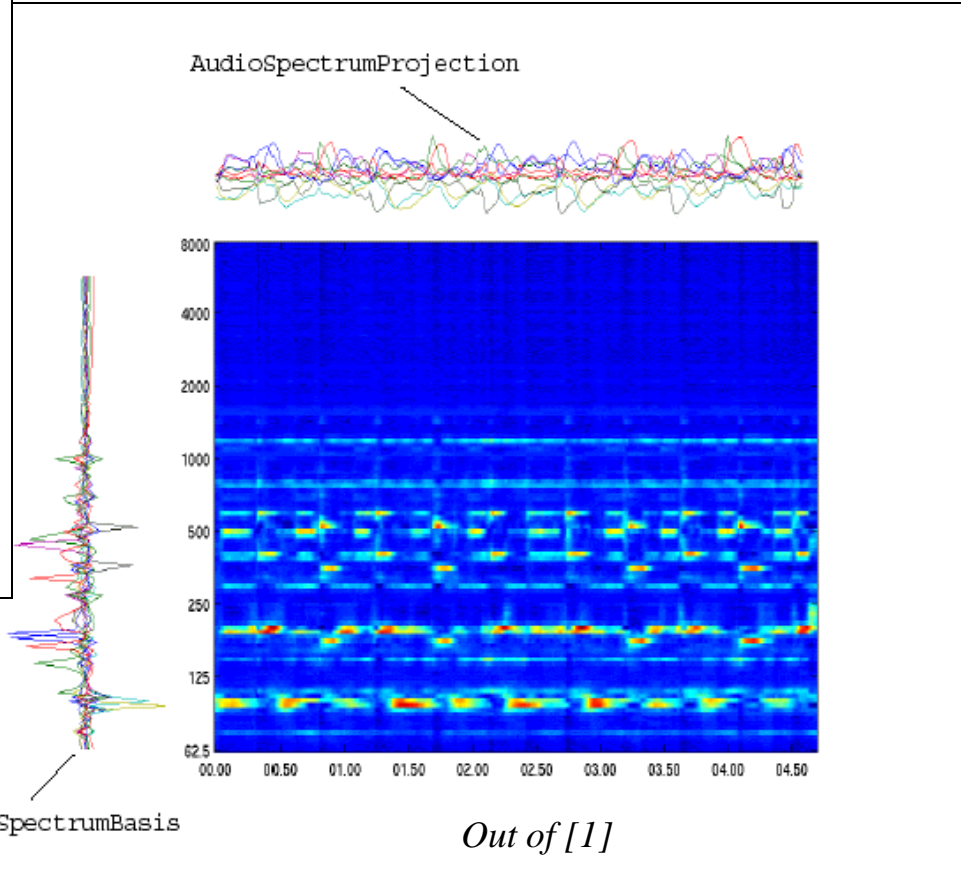
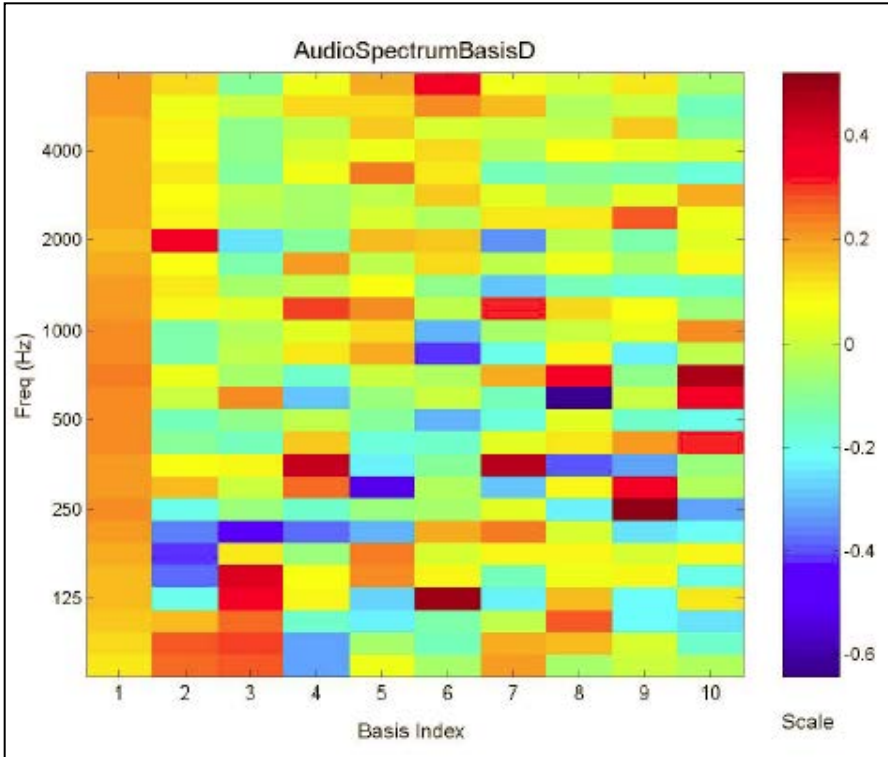
- **Timbral Spectral (Parametry widmowe)**

SpectralCentroid, HarmonicSpectralCentroid, HarmonicSpectralSpread, HarmonicSpectralVariation



- **Spectral Basis (Deskryptory dynamicznego widma sygnału)**

AudioSpectrumBasis, AudioSpectrumProjection

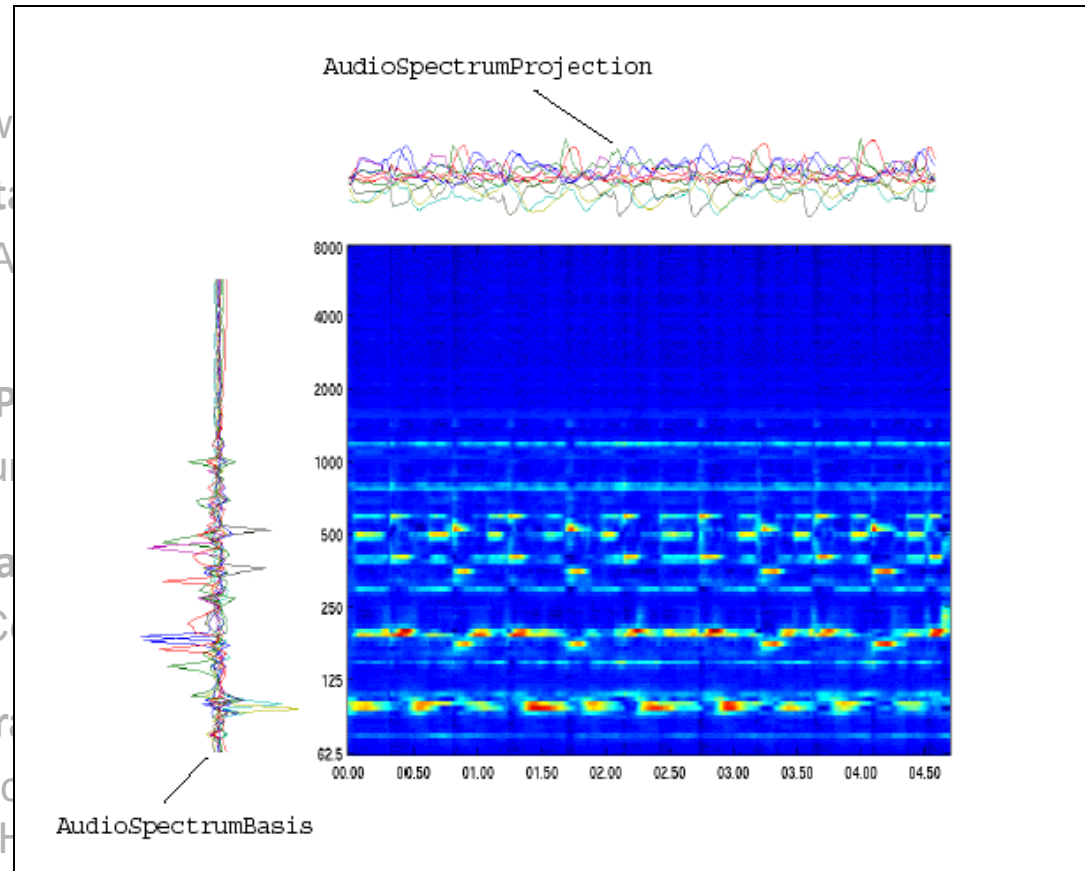


LogAttackTime, TemporalC

- **Timbral Spectral** (Par  
SpectralCentroid, Harmonic  
HarmonicSpectralSpread, H

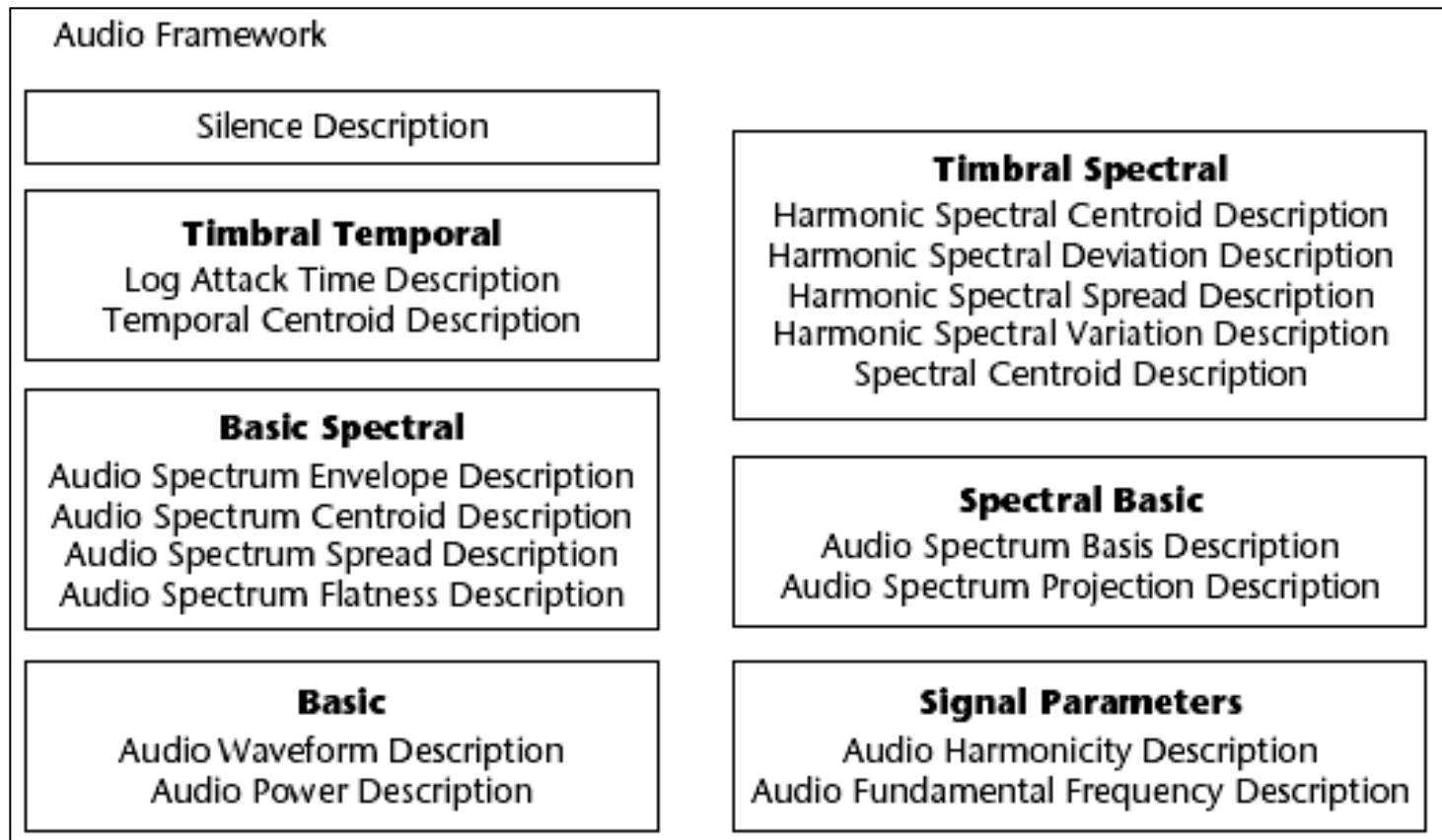
- **Spectral Basis** (Deskryptory dynamicznego widma sygnału)  
AudioSpectrumBasis, AudioSpectrumProjection

- **Basic (Podstawowe)**  
AudioWaveform, AudioPower
- **Basic Spectral (Podstawowe)**  
AudioSpectrumEnvelope, AudioSpectrumFlatness
- **Signal Parameters (Parametry)**  
AudioHarmonicicity, AudioFundamentalFrequency
- **Timbral Temporal (Parametry)**  
LogAttackTime, TemporalCentroid
- **Timbral Spectral (Parametry)**  
SpectralCentroid, Harmonicity, HarmonicSpectralSpread, HarmonicSpectralCentroid
- **Spectral Basis (Deskryptory dynamicznego widma sygnału)**  
AudioSpectrumBasis, AudioSpectrumProjection
- **Silence (Cisza)**



# MPEG-7

## GRUPY DESKRYPTORÓW AUDIO NISKIEGO POZIOMU



# Ekstrakcja cech dystynktywnych - parametryzacja

Wektor cech

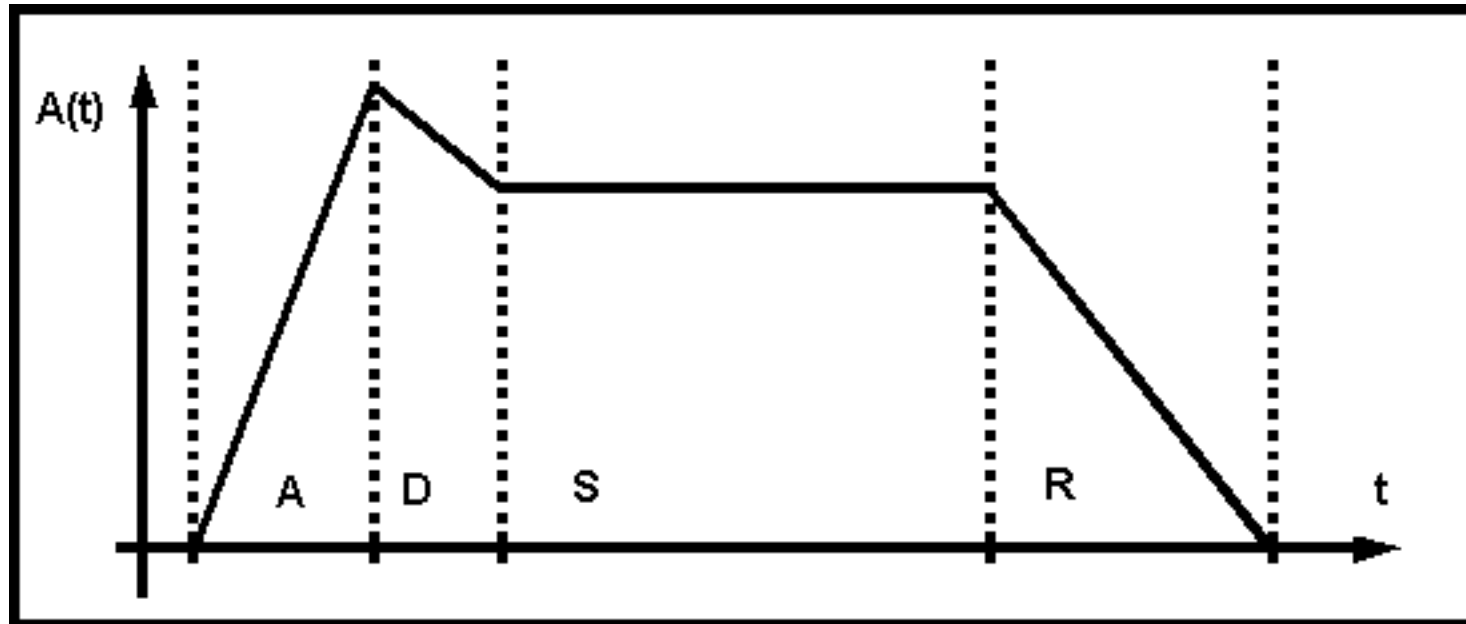
```
graph TD; A[Wektor cech] --> B[Parametry czasowe + statystyczne]; A --> C[Parametry widmowe + statystyczne]; A --> D[Parametry czasowo-częstotliwościowe];
```

**Parametry  
czasowe  
+  
statystyczne**

**Parametry  
widmowe  
+  
statystyczne**

**Parametry  
czasowo-  
częstotliwościowe**

# Model obwiedni ADSR



## Attack-Decay-Sustain-Release

- czas trwania: transjentu początkowego (ataku) i stanu quasi-ustalonego (ewentualnie znormalizowany względem okresu składowej podstawowej)
- szybkość wybrzmiewania [dB/oktawa]

- dewiacja częstotliwości  $n$ -tej składowej

$$\Delta f_n = \left[ (f_n - n \cdot f_1) / n \cdot f_1 \right] \times 100\%$$

- średnia ważona dewiacja częstotliwości dla grupy 5 najniższych harmoniczných

$$\overline{f_d} = \frac{\sum_{k=1}^5 A_k (\Delta f_k / (k f_1))}{\sum_{k=1}^5 A_k}$$

- $n_{fd}$  - numer składowej  $n_{fd}=1..5$   
o największej dewiacji

- parametry Tristimulus

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

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

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



- zawartość składowych parzystych w widmie

$$E_v = \sqrt{\sum_{k=1}^M A_{2k}^2} / \sqrt{\sum_{n=1}^N A_n^2} \quad M = \text{Entier}(N/2)$$

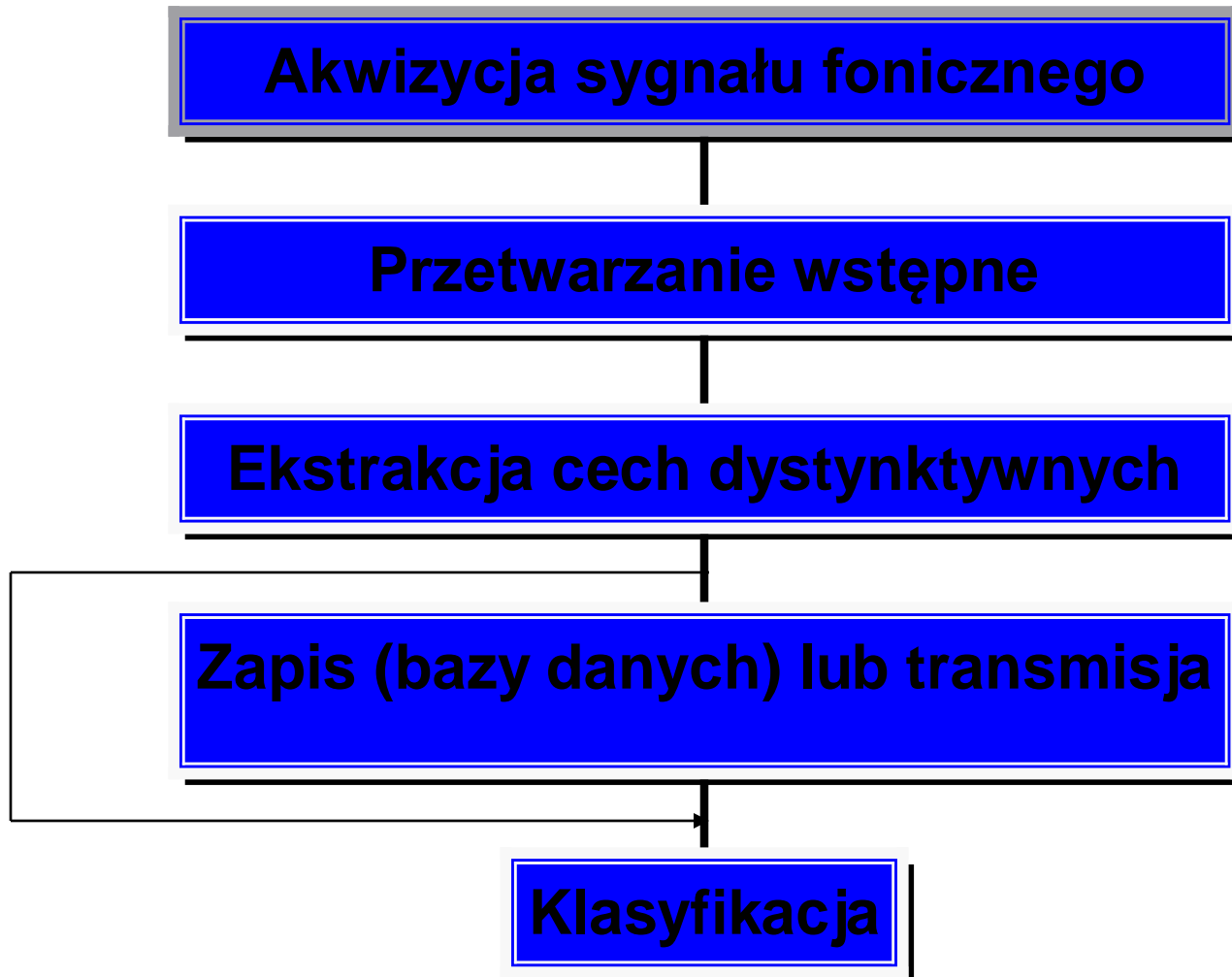
- zawartość składowych nieparzystych

$$O_{dd} = \sqrt{\sum_{k=2}^L A_{2k-1}^2} / \sqrt{\sum_{n=1}^N A_n^2} \quad L = \text{Entier}(N/2 + 1)$$

- jasność dźwięku

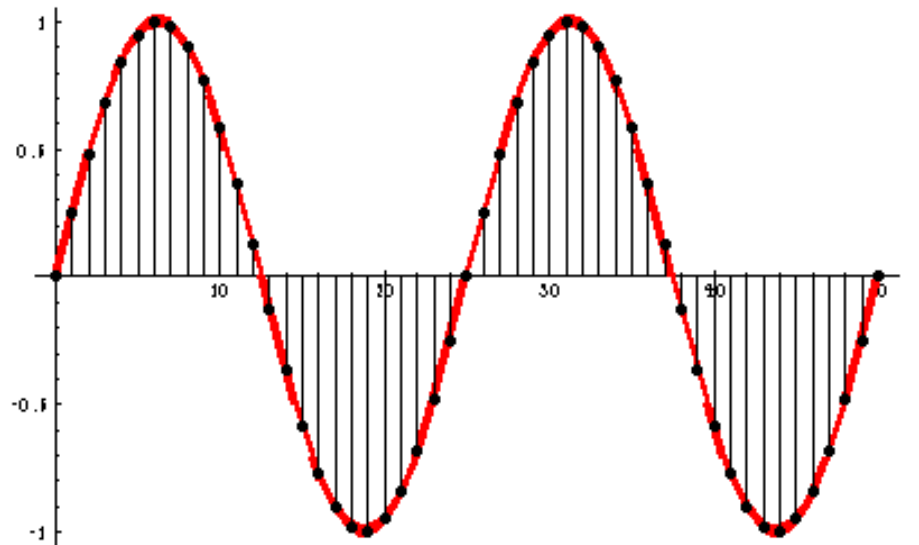
$$Br = \frac{\sum_{n=1}^N n \cdot A_n}{\sum_{n=1}^N A_n}$$

## Proces automatycznej klasyfikacji



# Akwizycja - reprezentacja sygnałów medycznych w systemach komputerowych

- sygnał rejestrowany przez czujniki pomiarowe jest zawsze sygnałem **ciągłym**
- Rejestrowany sygnał musi zostać poddany próbkowaniu

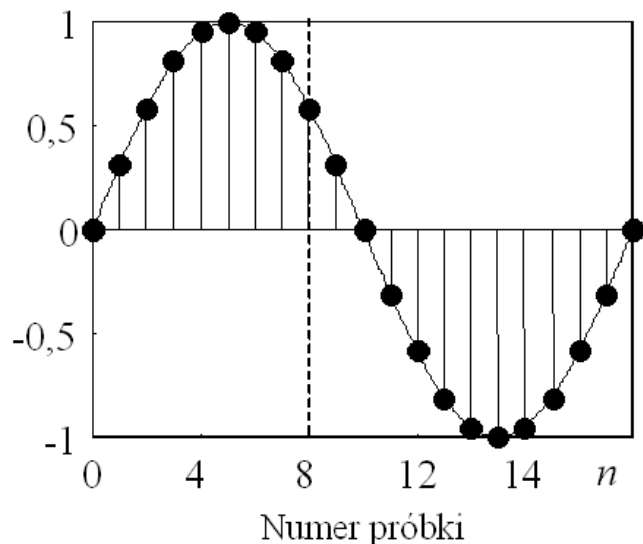


Przy wprowadzaniu do komputera ciągły sygnał z czujnika (ciągła czerwona linia na rysunku) zamieniany jest na szereg próbek w wybranych momentach czasu (czarne pionowe kreski)

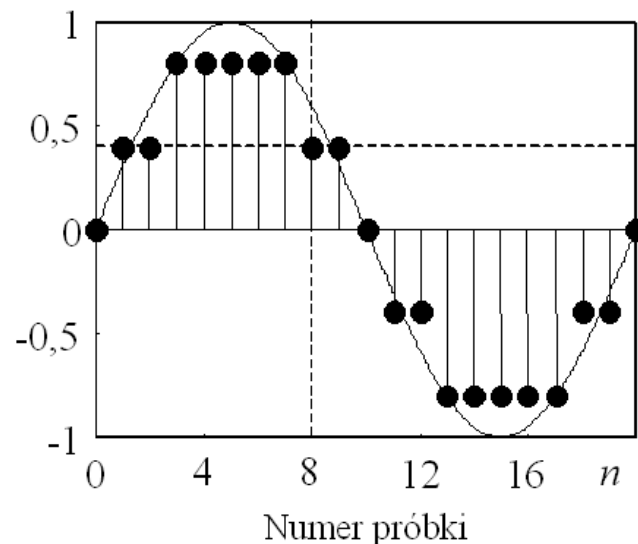
# Akwizycja - Reprezentacja sygnałów medycznych w systemach komputerowych

- Sygnał po próbkowaniu musi zostać poddany procesowi kwantyzacji

a)



b)



W pamięci komputera próbki są odwzorowywane z pewną ograniczoną dokładnością, co powoduje, że dozwolone są tylko niektóre wartości sygnału. Na rysunku ten efekt kwantowania przedstawiono w sposób przesadny, ale utrata dokładności następuje tu zawsze

## Akwizycja - Reprezentacja sygnałów medycznych w systemach komputerowych

- Oba procesy łącznie, to znaczy próbkowanie sygnału oraz jego kwantowanie dokonywane są w urządzeniu określonym jako przetwornik analogowo/cyfrowy. Dopiero tak spreparowany sygnał jest możliwy do umieszczenia w komputerze.

## Przetwarzanie wstępne

- Analiza czasowa (np. funkcja gęstości przejść przez zero)
- Transformacje widmowe: FFT, DFT
- Transformacja czasowo-częstotliwościowa:

gdzie:  $g_{b,a}(t)$  - funkcja analizująca,  
 $g(t)$  - funkcja prototypowa,  
 $a$  - współczynnik rozszerzenia,  
 $b$  - parametr przesunięcia czasowego

$$g_{b,a}(t) = \frac{1}{a} \cdot g\left(\frac{t-b}{a}\right)$$

- ciągła

- dyskretna

$$\hat{g}_{b,a}(\omega) = \sqrt{a} \cdot \hat{g}(a \cdot \omega) \cdot e^{j\omega b}$$

gdzie:  $\hat{g}(\omega)$  - transformata  
Fouriera funkcji  $g(t)$

$$DWT(a,n) = \frac{1}{\sqrt{a}} \cdot \sum_k h\left(\frac{k}{a} - n\right) \cdot x(k)$$

gdzie:  $k$  - indeks czasu,  
 $h(k)$  - funkcja prototypowa,  
 $x(k)$  - próbkowany sygnał poddawany analizie

- Analiza cepstralna:

gdzie:  $r$  - rząd współczynnika cepstralnego,

$l_{pr}$  - liczba próbek w ramce,

$i$  - numer kolejnej próbki widma,

$A_i$  - amplituda próbki  $i$ ,

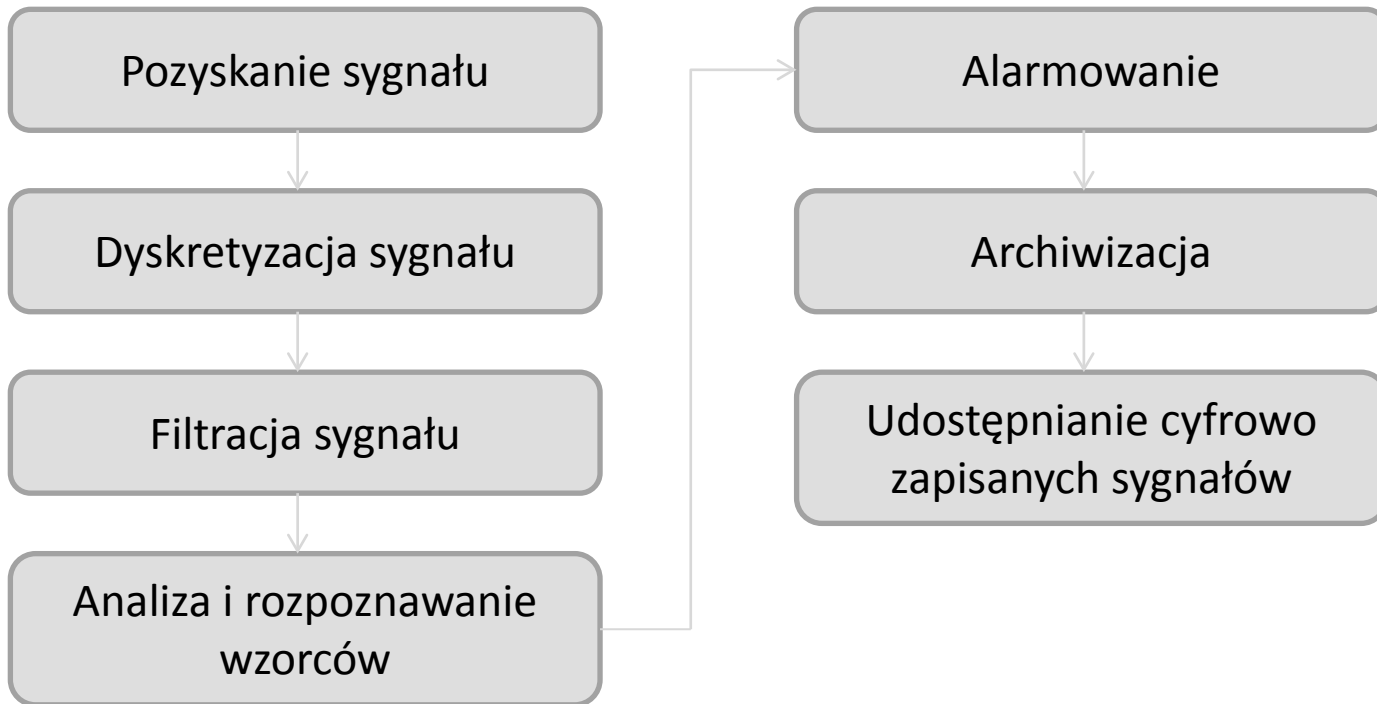
$f_p$  - częstotliwość próbkowania,

$f_c$  - maks. częstotl. w analizie cepstralnej

$$C_r = \sum_{i=1}^m \ln A_i \cdot \cos\left(r \cdot \frac{i \cdot \pi}{m}\right) \quad m = \frac{l_{pr}}{f_p} \cdot f_c$$

- Metody estymacji widma

# Proces pozyskiwania danych



# Przykład analizy korelacyjnej

$$r = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{X})^2 \sum_{i=1}^n (y_i - \bar{Y})^2}}$$

$r$  - współczynnik korelacji Pearsona

$\bar{X}, \bar{Y}$  - wartości średnie parameterów dla par instrumentów

$X, Y$



## Kryteria separowalności – metody topologiczne

- separowalność klas  $X, Y$  dla 2 parametrów - statystyka Fishera

$$|V| = \frac{\bar{X} - \bar{Y}}{\sqrt{S_1^2/k + S_2^2/m}}$$

- $\bar{X}, \bar{Y}$  - estymatory wartości średnich
- $S_1^2, S_2^2$  - estymatory wariancji
- separowalność klas bazy danych

$$Q = \min_{i,j} D_{i,j} / \max_i d_i$$

- $D_{i,j}$  - miara odstępów między klasami  $i, j$
- $d_i$  - miara rozrzutu obiektów w klasie  $i$

## Własności wybranych kryteriów separowalności

---

- statystyka Fishera:
  - ilustruje przydatność poszczególnych parametrów
  - im wyższa wartość  $|V|$ , tym większa przydatność parametru przy badaniu separowalności obiektów
  - nieprzydatna w przypadku dużych baz danych kryterium
- kryterium Q spełnione zadowalająco, gdy  $Q > 1$ 
  - zależy od doboru miar  $D_{ij}$  oraz  $d_i$
  - uzyskiwane miary zależą od stosowanej metryki