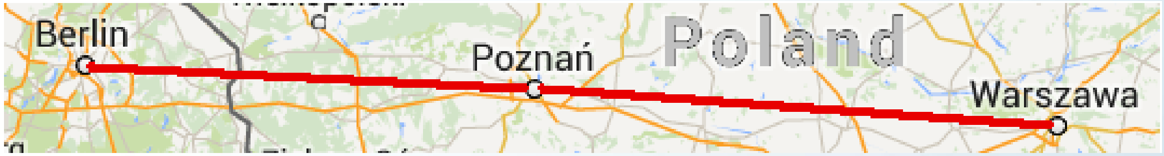
Content Based Audio Retrieval Methods. Cases of Violin Sound Modeling and Musical Genre Classification

Ewa Łukasik
Institute of Computing Science
Poznań University of Technology
Ewa.Lukasik@cs.put.poznan.pl

Outline

- Some self-marketing ☺
- Some audio basics
- Audio features
- Audio retrieval topics
- Musical genre classification
- Violin sound modeling
- Conclusions ?

Poznań



Poznań University of Technology - PUT

Institute of Computing Science

- One of 3 institutes in the Faculty of Computing.
- Its predecessor, the Institute of Control Engineering, was established in 1970 within the Faculty of Electrical Engineering.
- In 1988 it changed the name for the Institute of Computer Science, Automatics and Robotics,
- and finally (in 1990), after the extraction of a new chair, got its present shape.
- In years 2001-2010 was a division of the Faculty of Computer Science and Management.
- Since 2010 Institute of Computing Science is in the Faculty of Computing.
- The staff of the Institute consists of about 180 people, including over 100 persons occupying academic positions.

www2.cs.put.poznan.pl

CS Laboratories

- Laboratory of Operational Research and Artificial Intelligence
  - mobile systems research group
  - embedded systems research group
- Laboratory of Intelligent Decision Support Systems
- Laboratory of Computing Systems
  - Data Processing Technologies research group
  - Distributed Systems research group
- Laboratory of Algorithm Design and Programming Systems
  - bioinformatics,
  - software engineering and
  - scheduling research groups.

Journal

- The scope
  - approximate reasoning,
  - artificial intelligence,
  - combinatorial optimization,
  - computational complexity theory,
  - databases and data warehouses,
  - intelligent decision support,
  - knowledge engineering,
  - machine learning and data mining,
  - metaheuristics,
  - multiple criteria decision analysis,
  - networking and distributed systems,
  - parallel computing and concurrency,
  - production and project scheduling,
  - scheduling theory,
  - soft and granular computing,
  - software engineering.

Contacts Uni Siegen – PUT

UNIVERSITÄT SIEGEN

Studieninteressierte | Studierende | Beschäftigte (intern) | WissenschaftlerInnen | Alumni | Wirtschaft | Presse

/ start / news / forschung / 450727.html

- Startseite
- Universität
- Studium und Lehre
- Forschung
- Internationales
- Fakultäten
- Zentrum für Lehrerbildung
- Karriere
- **Aktuelles**
  - Baumaßnahmen
  - Stellenausschreibungen
  - News-Archiv
  - Amtliche Mitteilungen
- Kontakt

Kopernikus-Preis 2012 an Wissenschaftler aus Siegen und Poznan

Prof. Dr. Erwin Pesch

Professor Erwin Pesch, Wirtschaftsinformatiker der Universität Siegen, und Professor Jacek Błażewicz, Informatiker aus Poznan, erhalten für ihre Verdienste um die deutsch-polnische Zusammenarbeit in der Wissenschaft den Kopernikus-Preis der Deutschen Forschungsgemeinschaft (DFG) und der Stiftung für die polnische Wissenschaft (FNP).

In der Begründung der Jury heißt es: „Die beiden Wissenschaftler leisten hervorragende Forschungsarbeit, engagieren sich für den Nachwuchs, kooperieren seit vielen Jahren und ergänzen sich dabei auf optimale Weise." Der Kopernikus-Preis ist mit 100 000 Euro dotiert und nach dem Leibniz-Preis der höchstdotierte Forschungspreis der DFG. Er wird am 17. September 2012 in Warschau von den Präsidenten der DFG und FNP, Professor Matthias Kleiner und Professor Maciej Żylicz, verliehen.

Contacts Uni Siegen – PUT

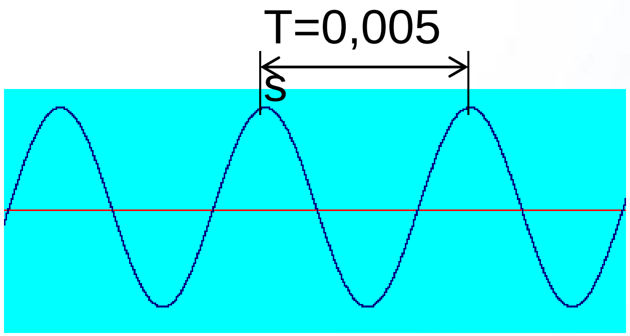
Verleihung des akademischen Grades Doktor der Wirtschaftswissenschaften ehrenhalber

(21.06.2006) Verleihung des akademischen Grades Doktor der Wirtschaftswissenschaften ehrenhalber (Dr. rer. pol. h.c.) an Herrn Univ.-Prof. Dr. **Jacek Blazewicz.** In Anerkennung seiner hervorragenden wissenschaftlichen Leistungen in der Betriebswirtschaftslehre, Wirtschaftsinformatik und Informatik, insbesondere in der Produktion, Projektmanagement und Scheduling.

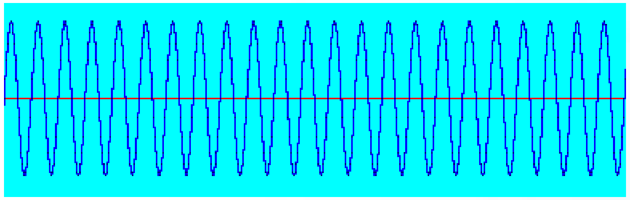
Back to main subject…

- Some audio basics….

Simple tones



200 periods / 1 sek = **200 Hz**



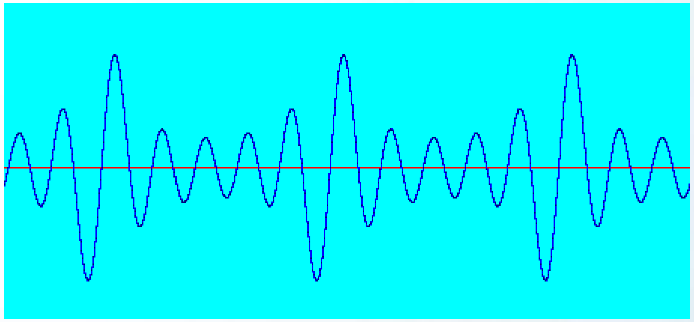
2000 periods / 1 sek = **2000 Hz**
(2 kHz)

Range of audible sounds (**Hz –> Hertz**)

...... **20 Hz** ...... **200 Hz** ...... **2 kHz** ...... **20 kHz** ......

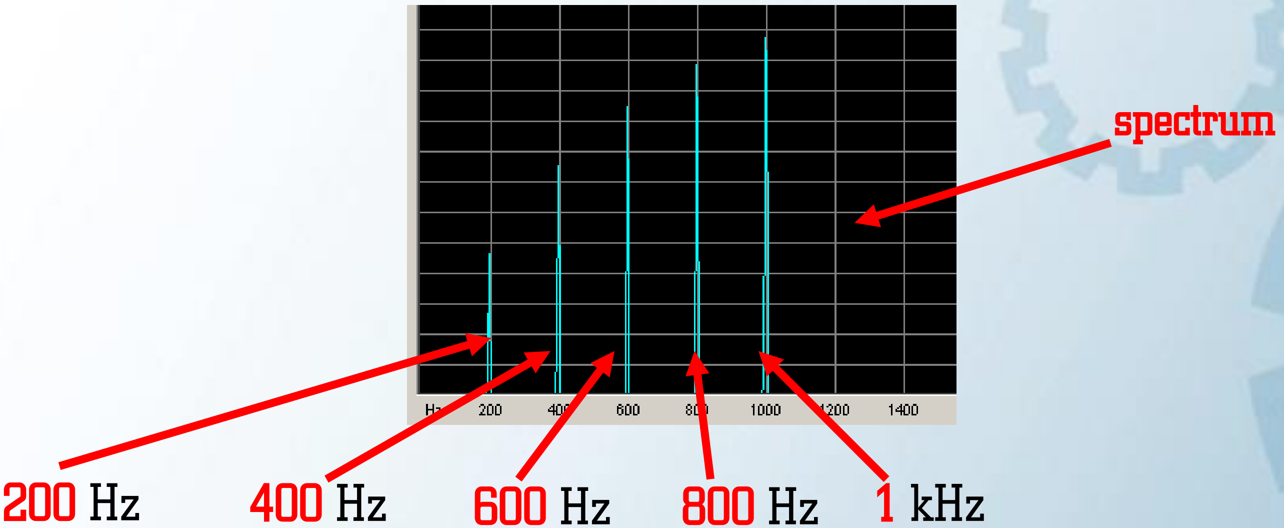
Low frequency ***high frequency***

More complex sound and its spectrum

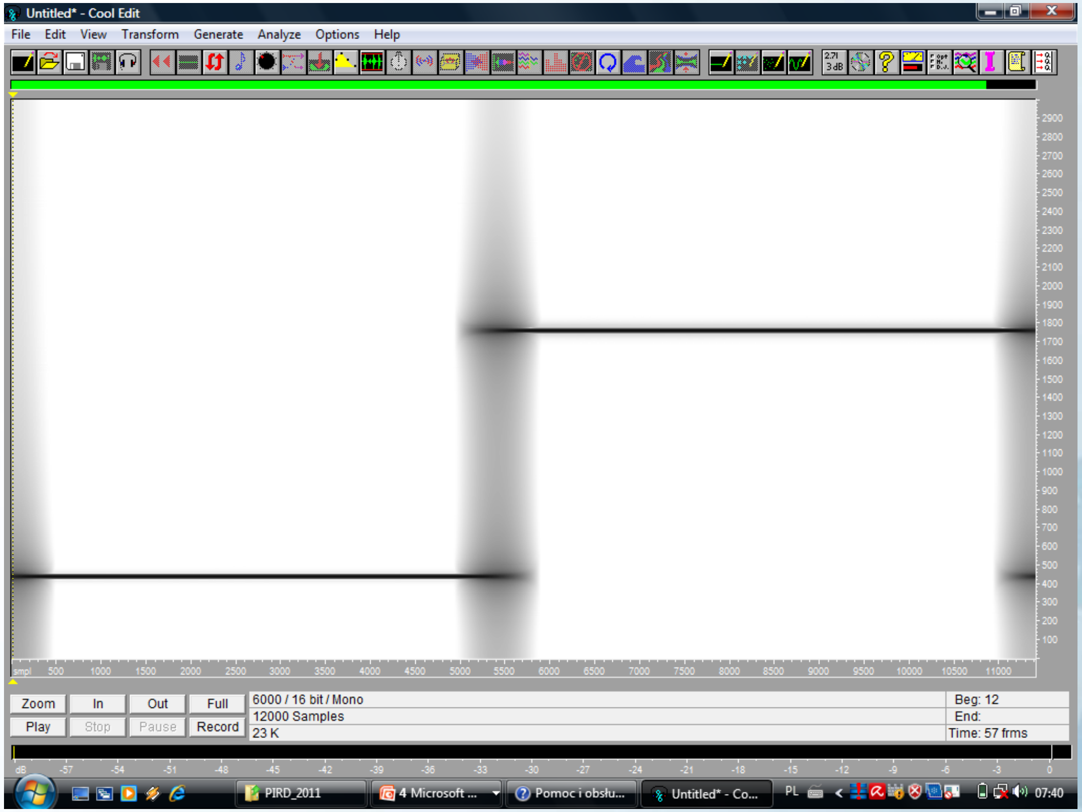


The above signal was built of tones of 200Hz , 400Hz, 600Hz , 800Hz and 1 kHz

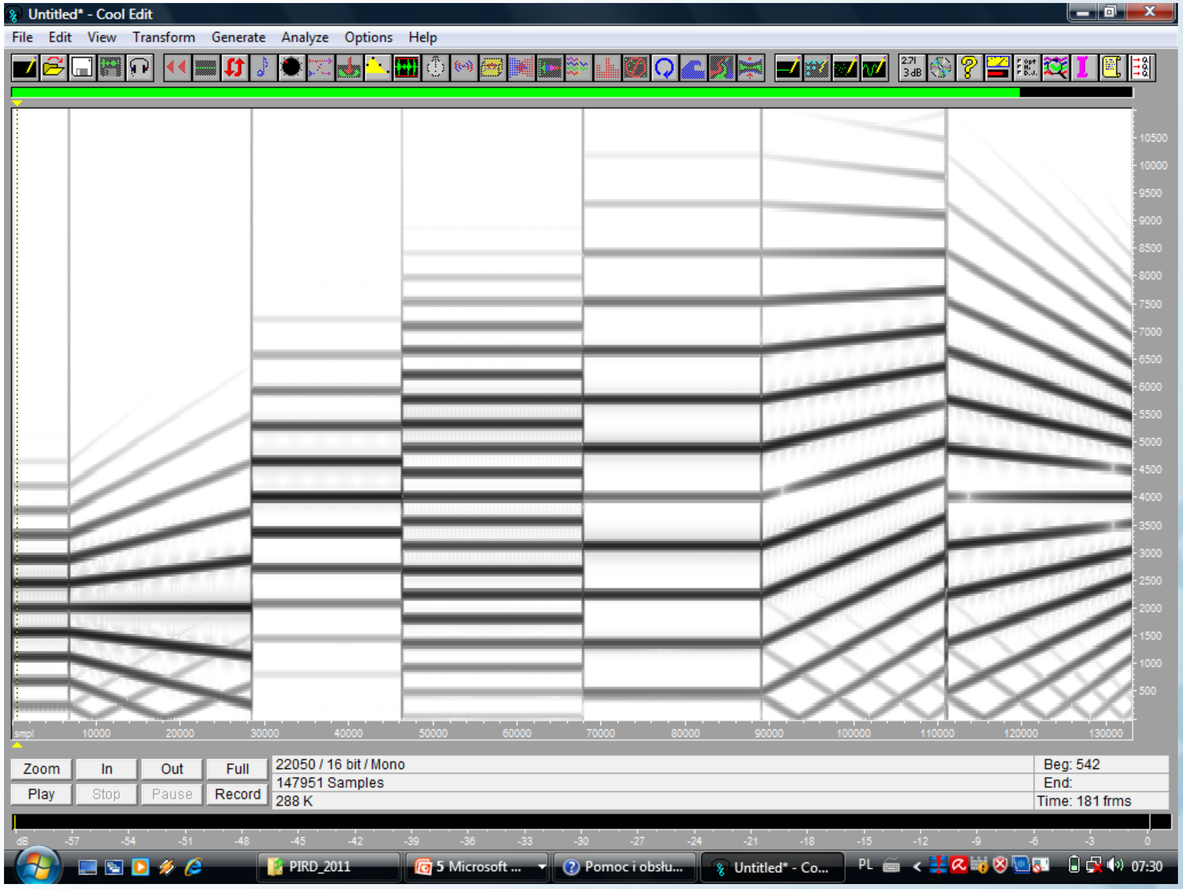
Is there a way **to see** the frequencies and amplitudes of this sound?



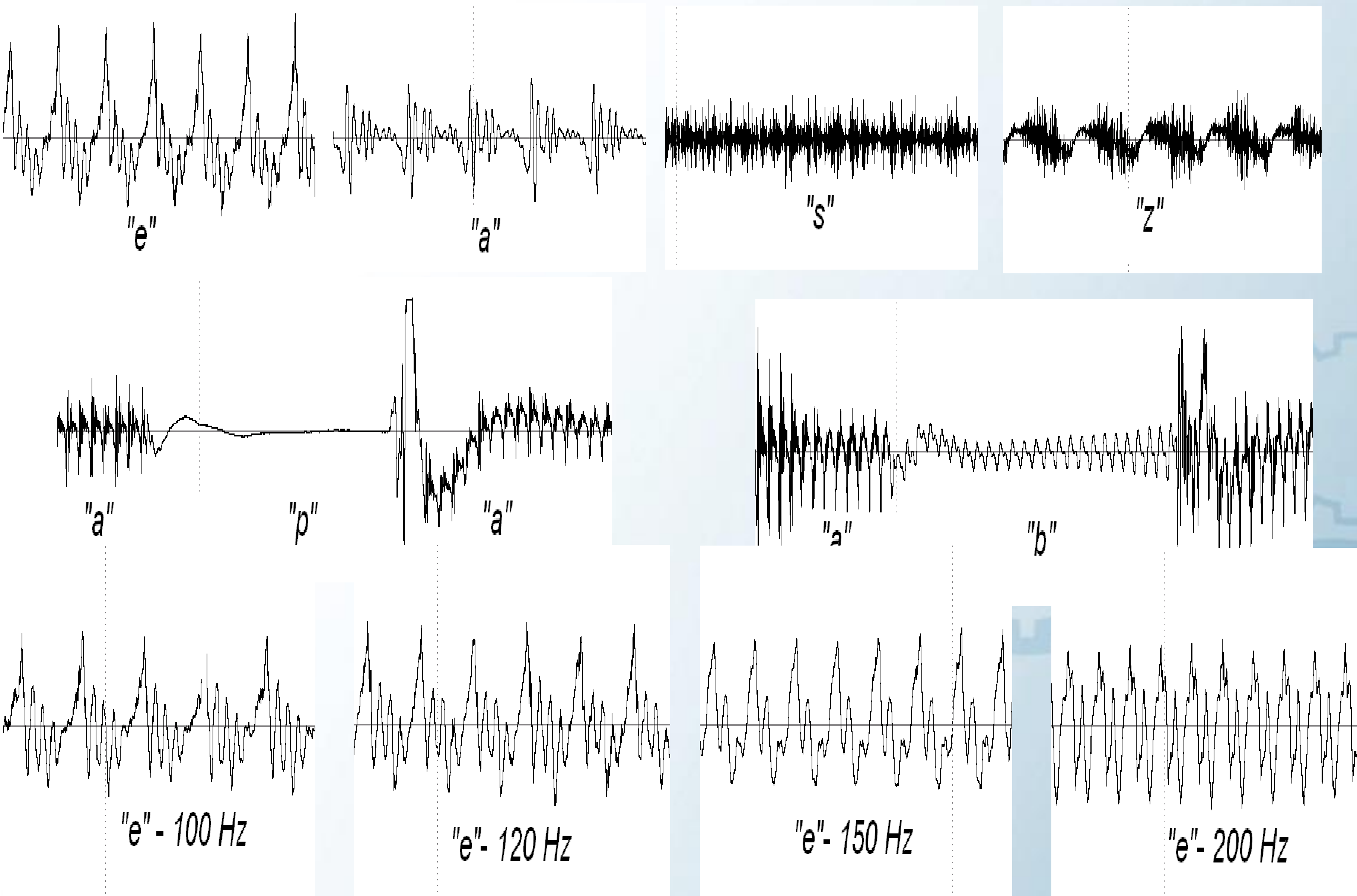
Spectrogram



Some other examples

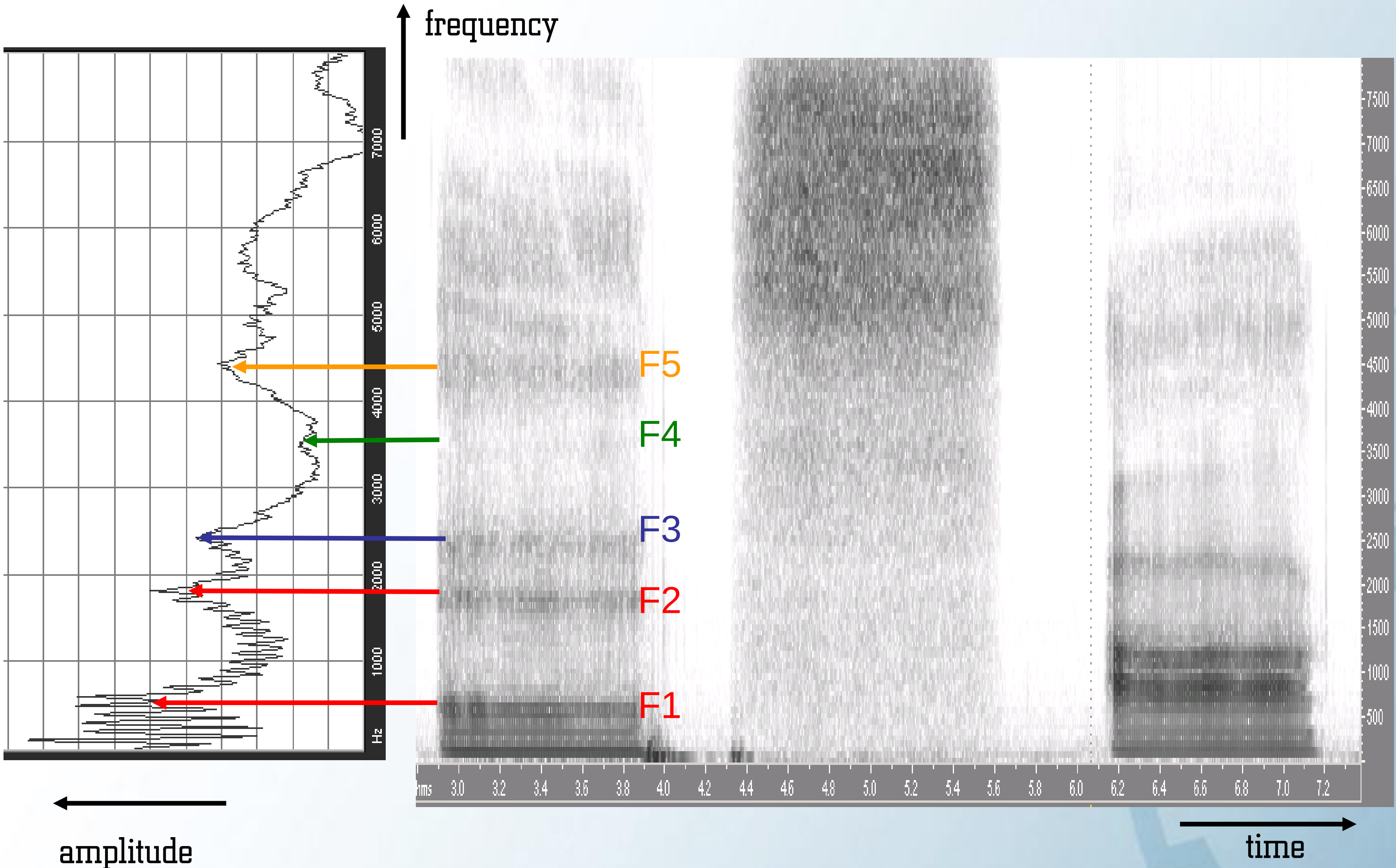


Waveforms of speech phonemes

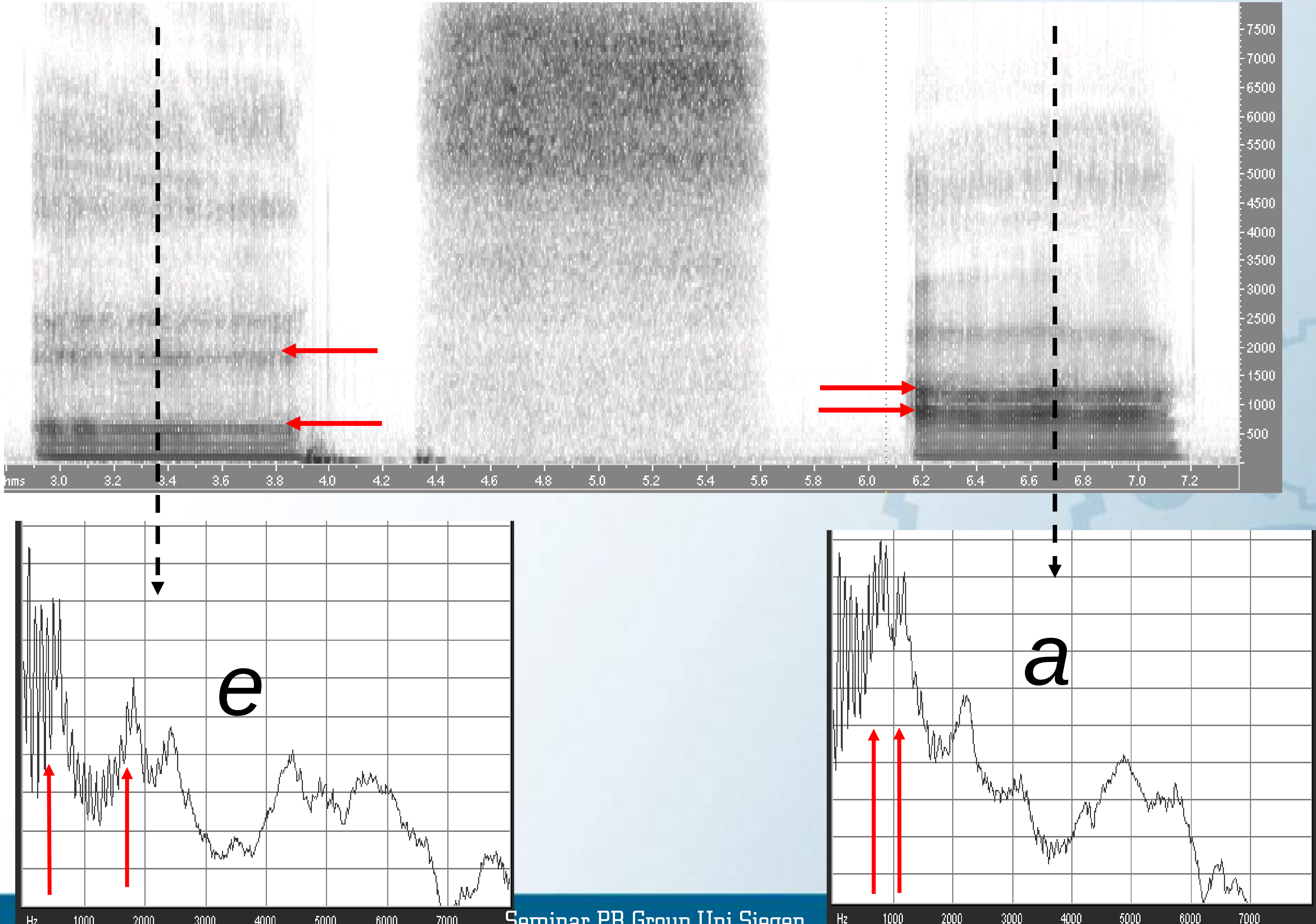


spectrum

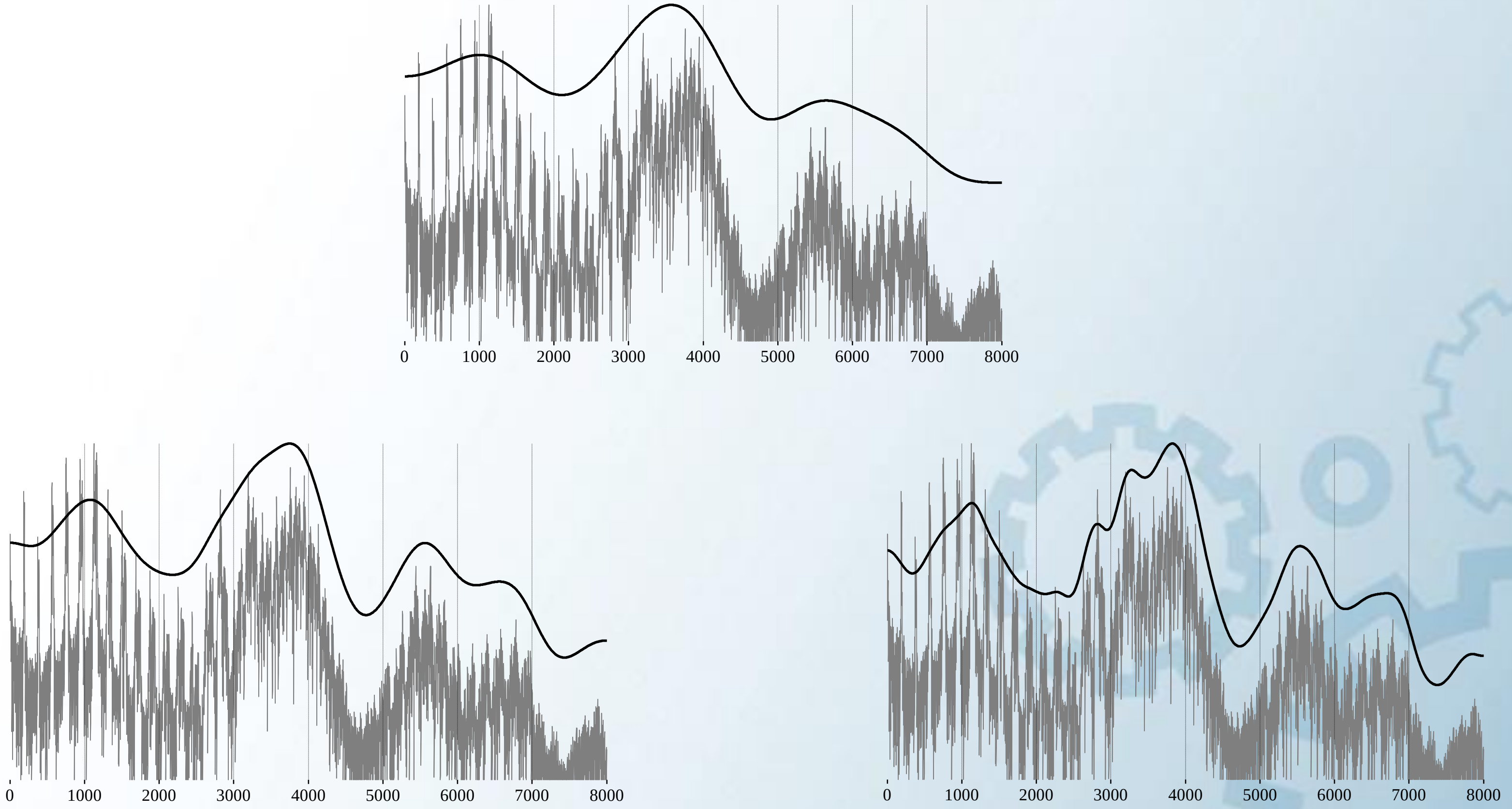
Spectrogram of speech phonemes



What are the differences between phonemes in spectrum or spectrogram ?



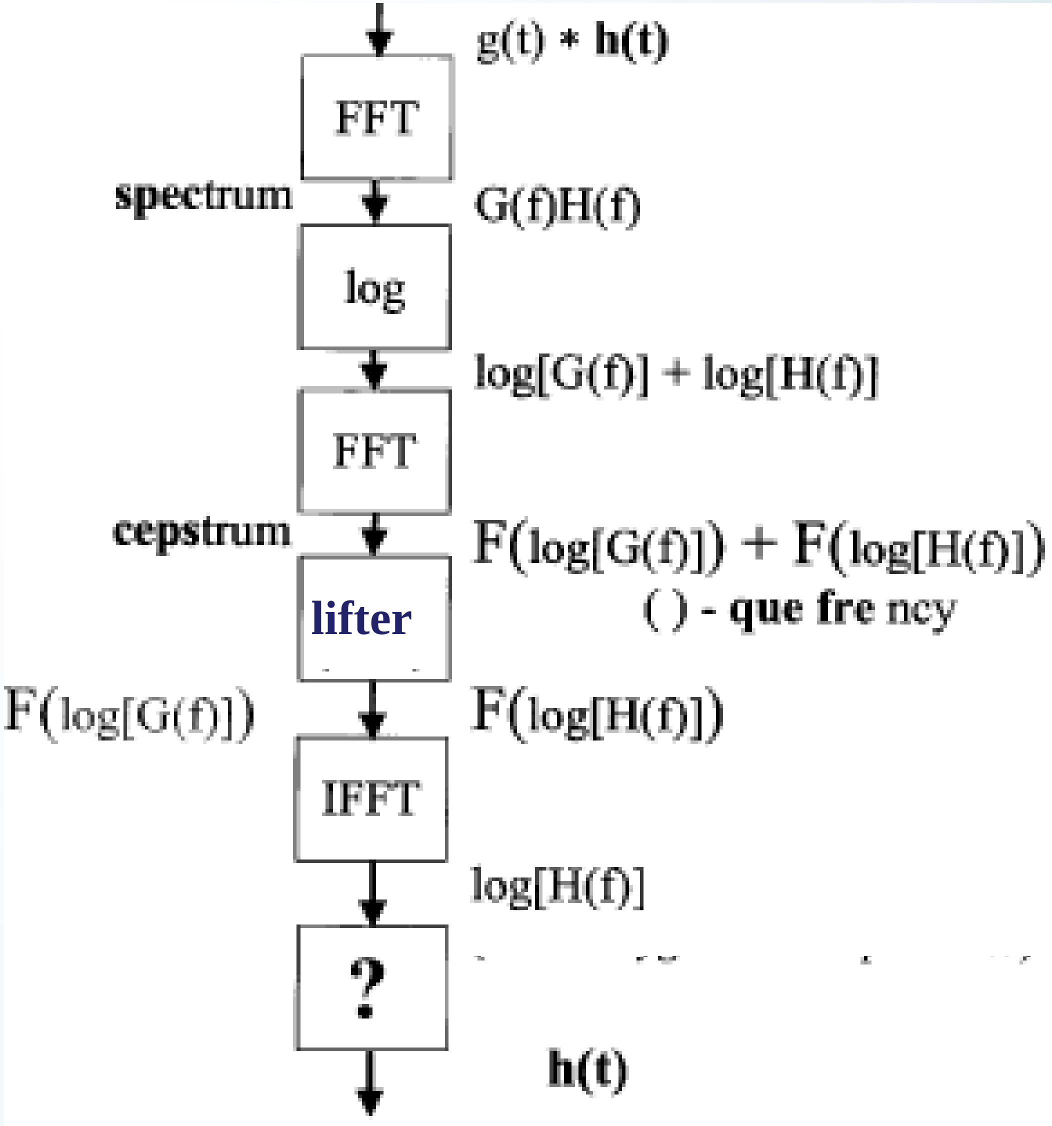
Spectral envelope



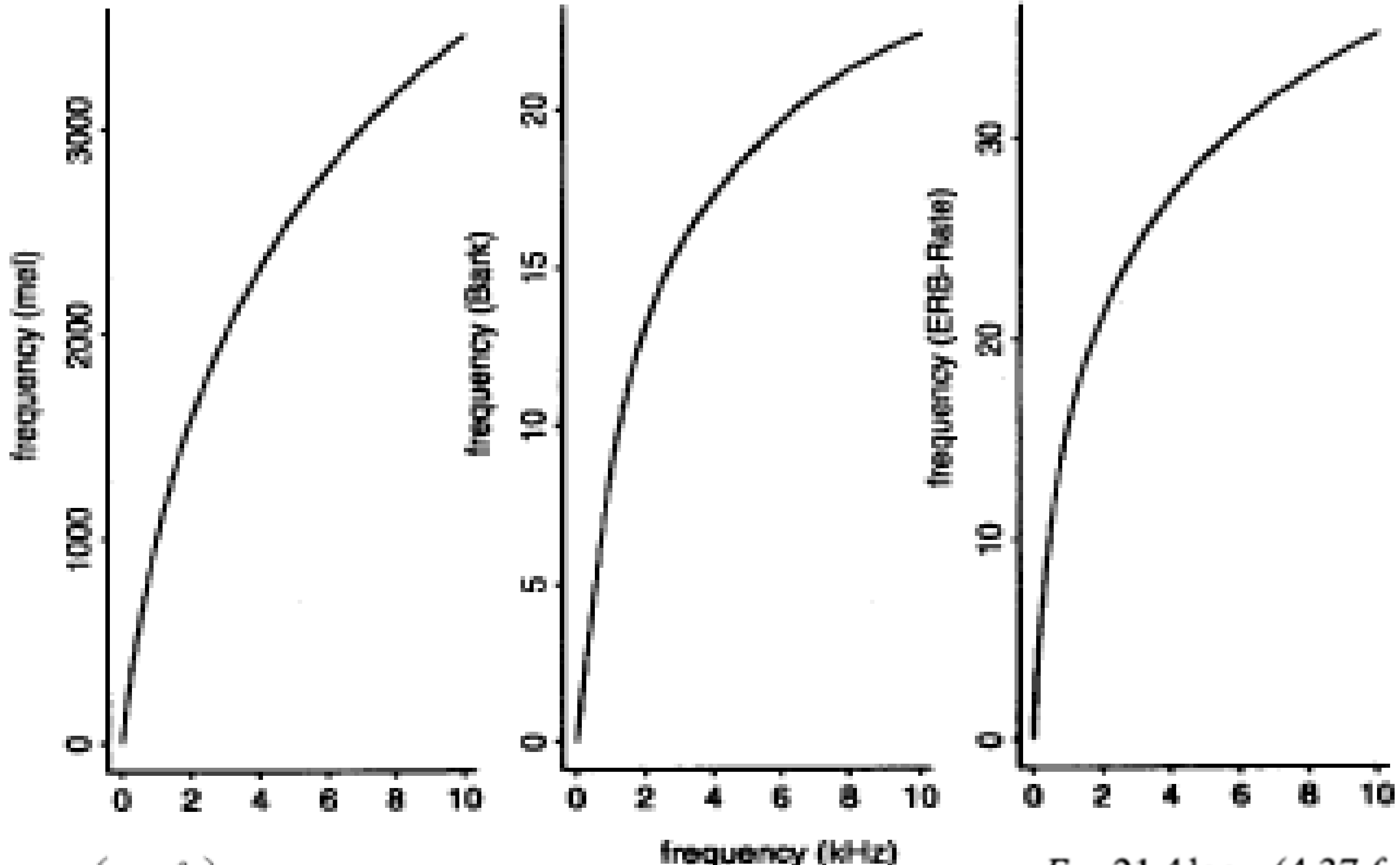
How to calculate the spectral envelope?

Several methods….

Cepstral analysis



Perceptual scales



$$Mel(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

$$f_{Bark} = 13 tan^{-1}(0.0076 f_{Hz}) + 3.5 tan^{-1}\left(\frac{f_{Hz}}{7.500}\right)$$

$$E = 21.4 \log_{10}(4.37 f_{kHz} + 1)$$

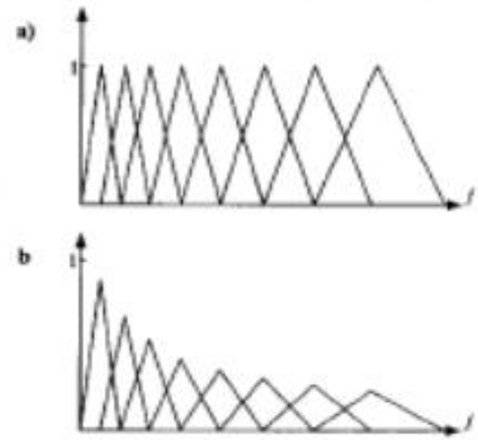
MFCC – Mel Frequency Cepstral Coefficents

The most often
used feature
for audio description

$$s'_n = s_n - k s_{n-1}$$

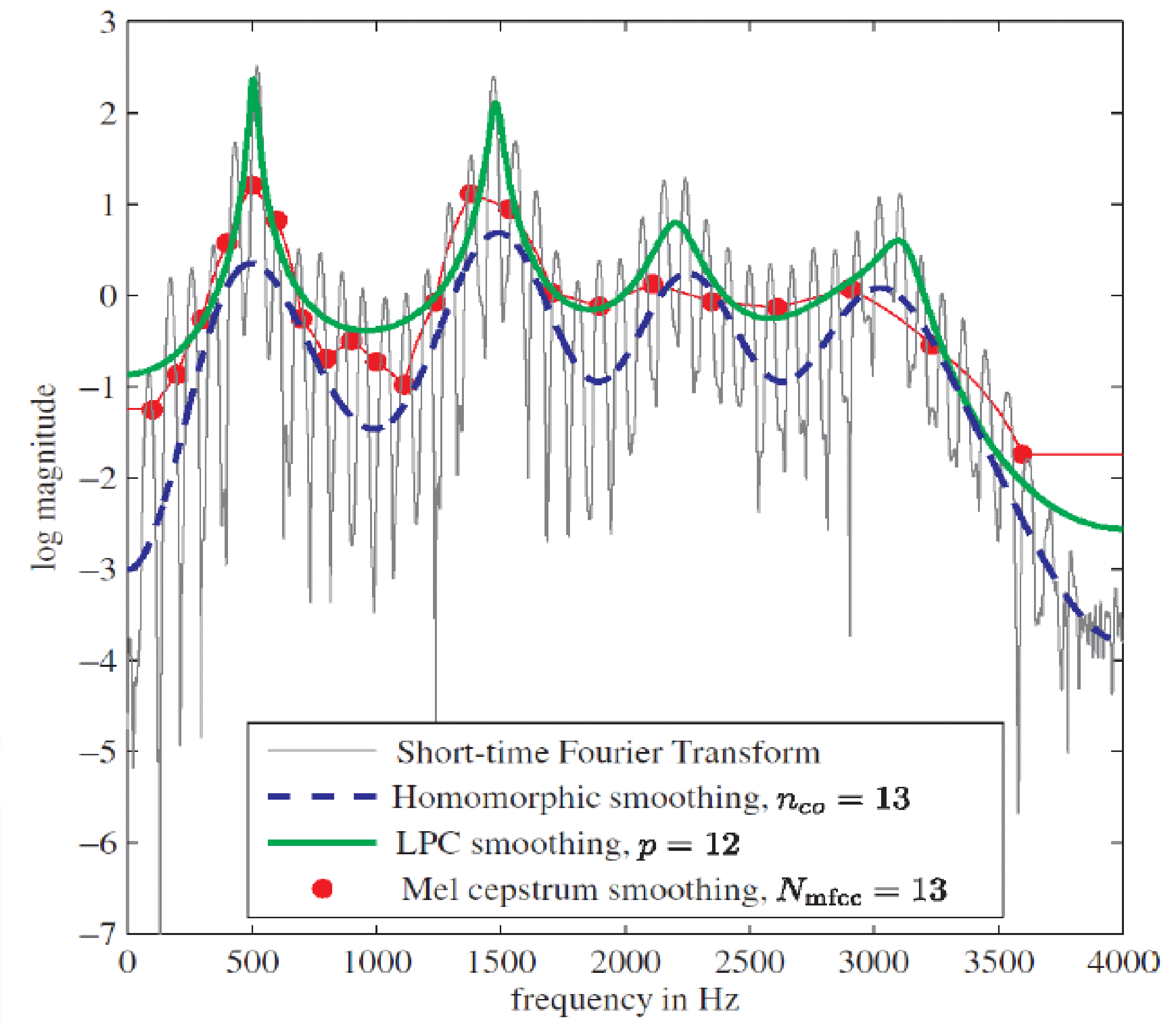
$$s'_n = \left[0.54 - 0.46\cos\left(\frac{2\pi(n-1)}{N-1}\right)\right] s_n$$

$$Mel(f) = 2595\log_{10}\left(1+\frac{f}{700}\right)$$



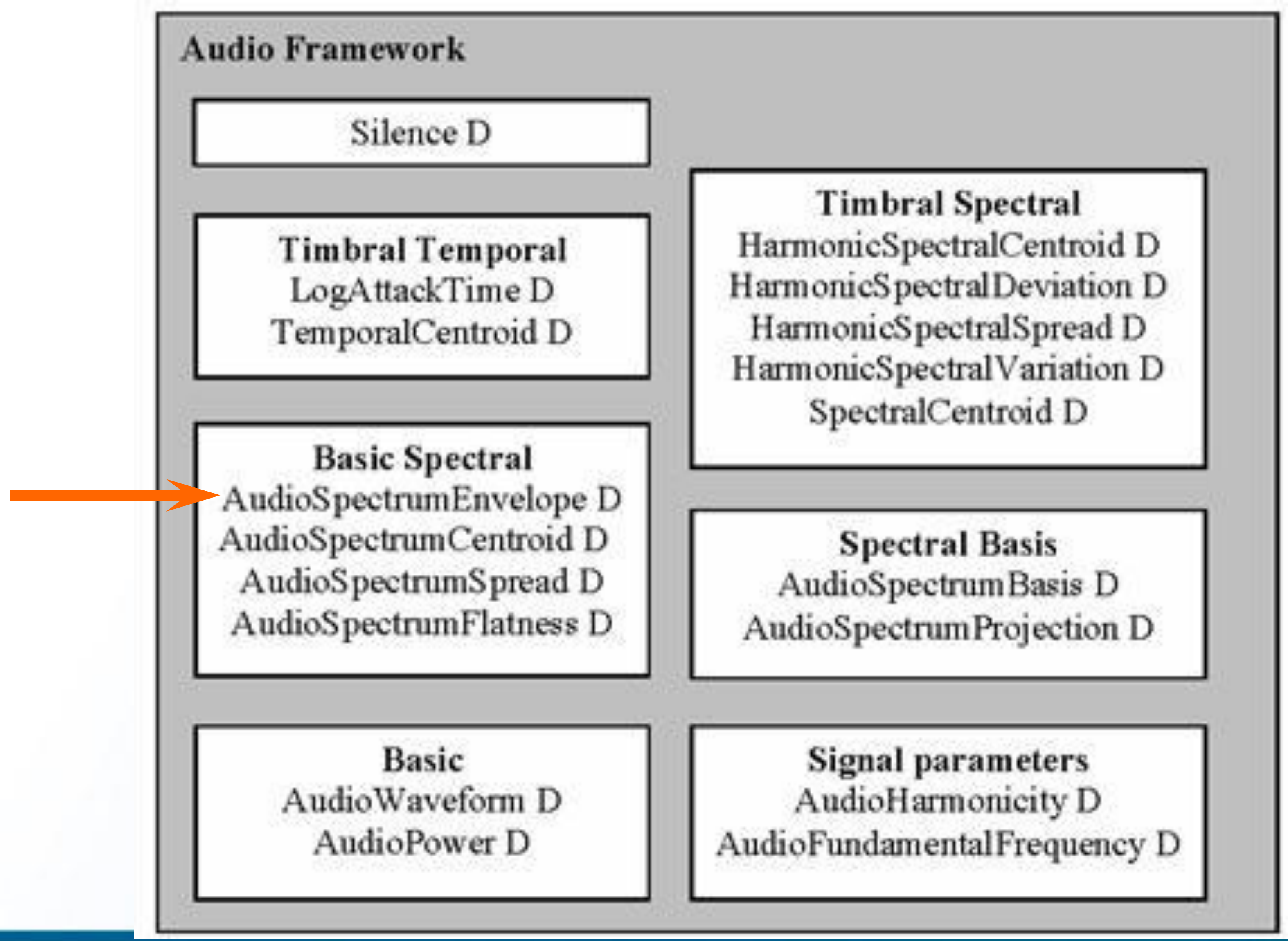
$$c_n = \sqrt{\frac{2}{N}}\sum_{i=1}^{N} m_i \cos\left(\frac{\pi n}{N}(i-0.5)\right)$$

Comparison of methods for spectrum enveloping



Source: Rabiner & Schafer, 2007

Yet another one… MPEG-7



Exemplary MPEG-7 timbral spectral descriptors

harmonic spectral centroid

$$HSC = \frac{\sum_{h=1}^{nb\_h} f(h) \cdot A(h)}{\sum_{h=1}^{nb\_h} A(h)}$$

harmonic spectral spread

$$HSS = \frac{1}{HSC} \sqrt{\frac{\sum_{h=1}^{nb\_h} A^2(h) \left(f(h) - HSC \right)^2}{\sum_{h=1}^{nb\_h} A(h)}}$$

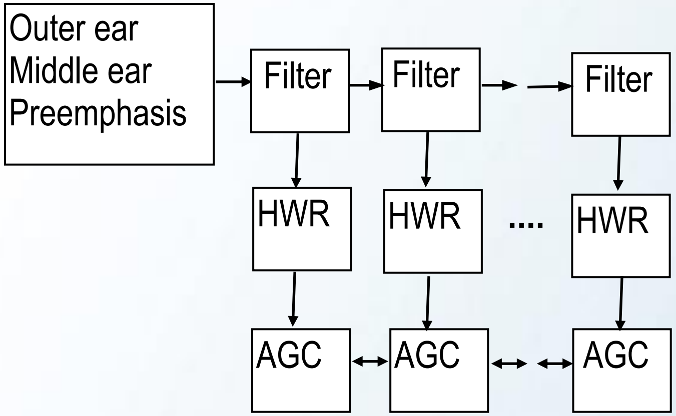
harmonic spectral deviation

$$HSD = \frac{\sum_{h=1}^{nb\_h} \left| \log_{10} A(h) - \log_{10} SE(h) \right|}{\sum_{h=1}^{nb\_h} \log_{10} A(h)}$$

harmonic spectral variation

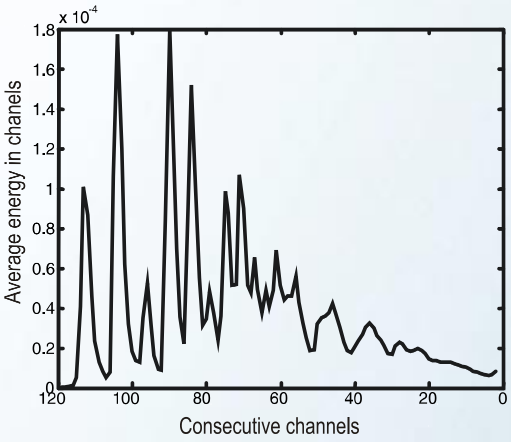
$$HSC = \frac{\sum_{h=1}^{nb\_h} A_{-1}(h) \cdot A(h)}{\sqrt{\sum_{h=1}^{nb\_h} A_{-1}^2(h)} \cdot \sqrt{\sum_{h=1}^{nb\_h} A^2(h)}}$$

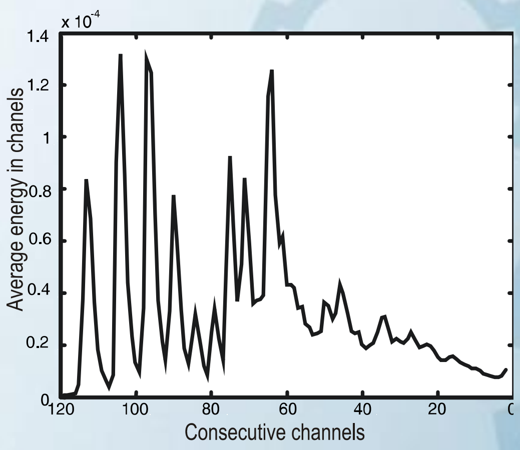
Lyon's cochlear model filters



half wave rectifiers

automatic gain control





Feature extraction

- To extract suitable features that capture relevant key aspects while suppressing irrelevant details or variations.
- The notion of *similarity* is of crutial importace in the design of audio features
- In certain applications one may be interested in characterizing audio recording irrespectively of certain details concerning the interpretation or instrumentation.
- Other applications may be concerned with measuring just the subtleties that relate to a musician's individual articulation, emotional expressiveness or features.
- Most applied approach: **the bag of features**
- From low-level description through higher level description to semantic description -> computing with words

Audio retrieval topics

- automatic speech recognition,
- audio forensics,
- music information retrieval, MIR,
- music emotion retrieval, MER,
- audio segmentation,
- environmental sound retrieval,
- audio surveillance,
- audio based/supported video retrieval,
- …. many others.

Music Information Retrieval

ISMIR – flagship conference – yearly since 2000

MIREX competition

- Audio Train/Test Tasks
  - Audio Artist Identification
  - **Audio Genre Classification**
  - Audio Music Mood Classification
  - Audio Classical Composer Identification
- Symbolic Genre Classification
- Audio Onset Detection
- Audio Key Detection
- Symbolic Key Detection
- Audio Tag Classification
- Audio Cover Song Identification
- Real-time Audio to Score Alignment (a.k.a Score Following)
- Query by Singing/Humming
- Audio Melody Extraction
- Multiple Fundamental Frequency Estimation & Tracking
- Audio Chord Estimation
- Query by Tapping
- Audio Beat Tracking
- Structural Segmentation
- Audio Tempo Estimation
- Discovery of Repeated Themes & Sections
- Audio Downbeat Estimation
- Audio Fingerprinting
- Singing Voice Separation
- **Grand Challenge on User Experience**

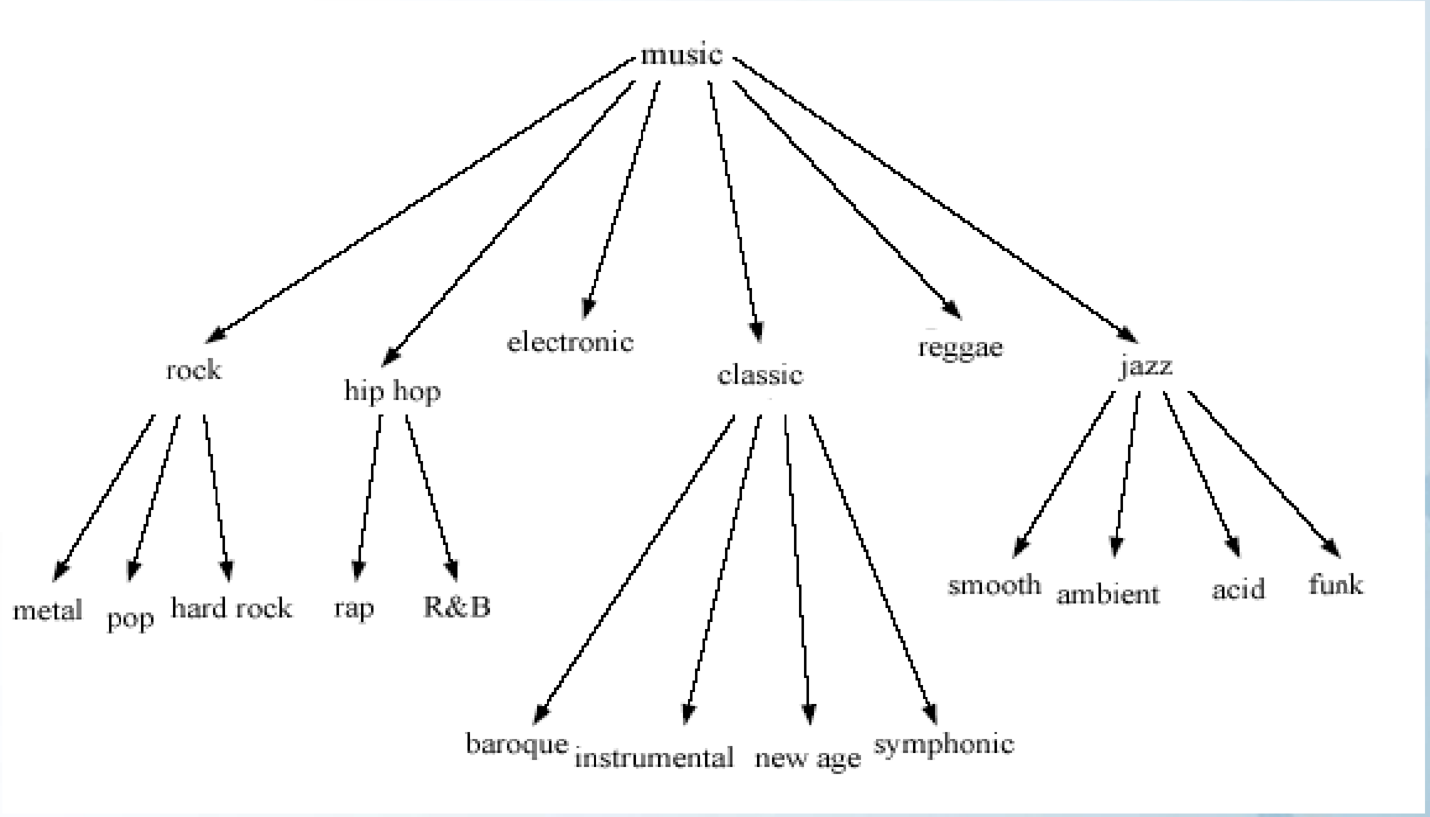
http://www.music-ir.org/mirex/wiki/

More advanced topics of MIR

- Recognition of musical instuments within the mixes
- Sound source separation in polyphony
- Singing speech recognition

Music genre classification

- Hierarchical taxonomy (tracks from *Jamendo.com*)



Ł. Dropik, E. Łukasik, Two-Level Hierarchical Classification of Music Genre for Music Social Networks, Foundations of Computer and Decision Sciences, Vol. 35, No. 4, 2010, pp. 261-282

Features

- MFCC
- Spectral Flux

$$SF(n,l) = \sum_{l=0}^{N-1} X(n,l) - X(n-1,l)$$

X – spectrum, n – spectrum frame index, l– frequency bin index

- Zero Crossing Rate

$$ZCR_k = \frac{1}{2}\sum_{i=1}^{m} \left| sign\ \ x(i) \right| - \left| sign\ \ x(i-1) \right|$$

x(i) - time domain sample in the time frame k, m – time frame length

- Average silence ratio (legato vs. staccato)

$$ASR = \frac{1}{2N}\sum_{n=0}^{N-1} 1 - sgn(STE(n) - \rho \cdot avgSTE\ \ )$$

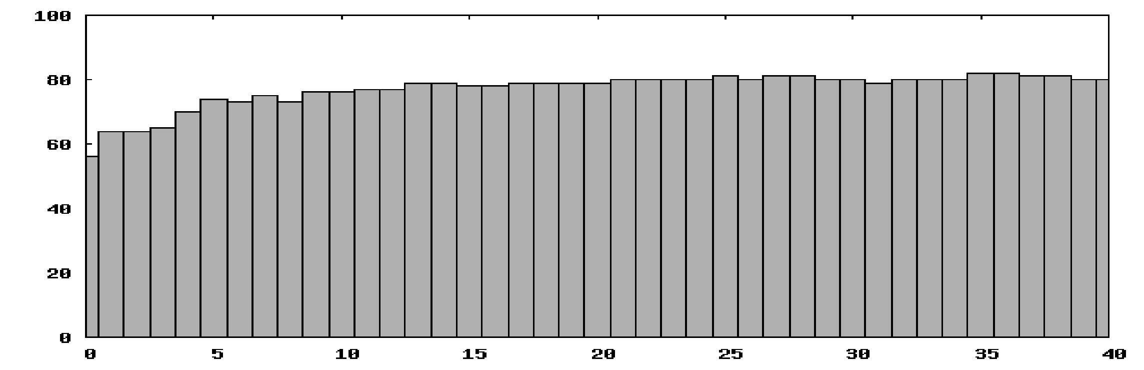
$$STE(n) = \sum_{i=n \cdot m}^{(n+1)m-1} x^2(i) \qquad avgSTE = \frac{1}{N}\sum_{n=0}^{N-1} STE(n)$$

N – number of frames in the window, , m – time frame length, ρ - silence threshold (experimental)

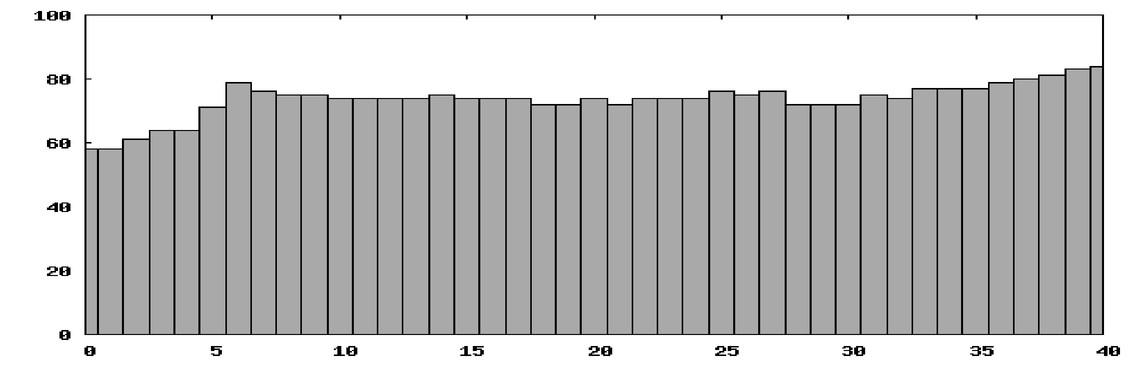
- Tempo histogram
  - Quantized between 60 and 280 bpm (autocorrelation)

Number of MFCCs vs. classification rate

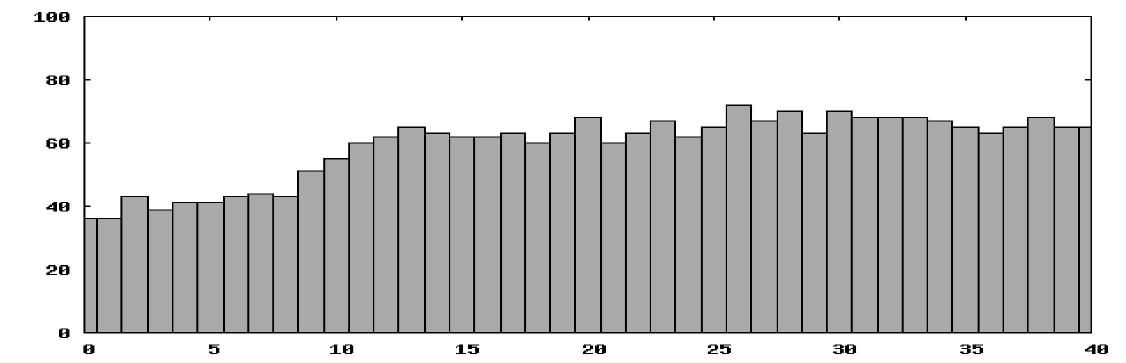
1st level of Taxonomy/SVM



Classical/SVM

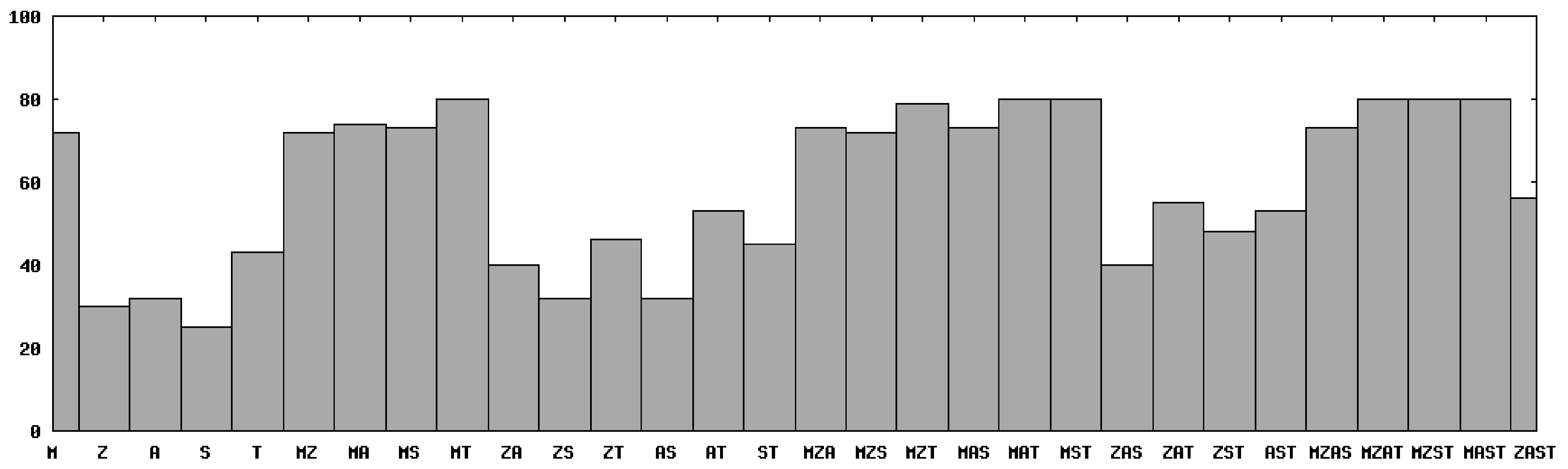


Jazz/SVM

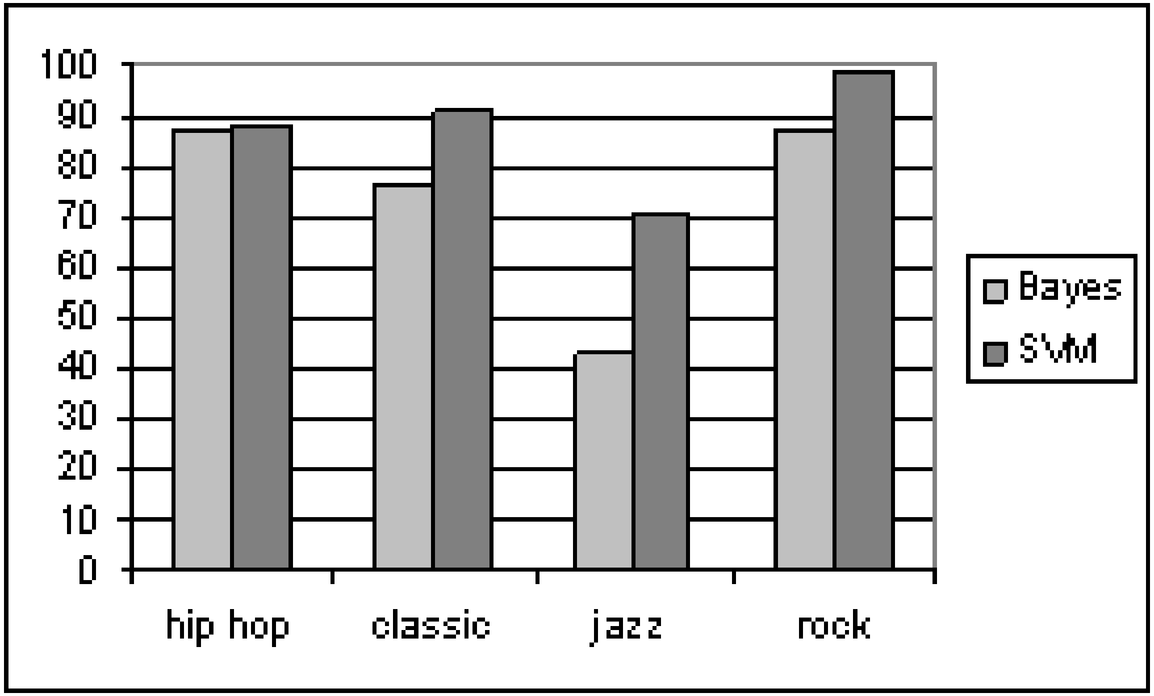


Features combination vs. classification rate

- 1st level of Taxonomy



Classification results



ISMIS 2011 contest

- KOSTEK B., KUPRYJANOW A., ŻWAN P., JIANG W., RAŚ Z., WOJNARSKI M., ŚWIETLICKA J. (2011), Report of the ISMIS 2011 Contest: Music Information Retrieval, Foundations of Intelligent Systems, ISMIS 2011, Springer Verlag, 715–724, Berlin, Heidelberg
- A database of 60 music musicians/performers was prepared for the competition.
- The material is divided into six categories:
  - classical music
  - jazz,
  - blues,
  - rock,
  - heavy metal,
  - pop

http://tunedit.org/challenge/music-retrieval

ISMIS 2011 contest Features

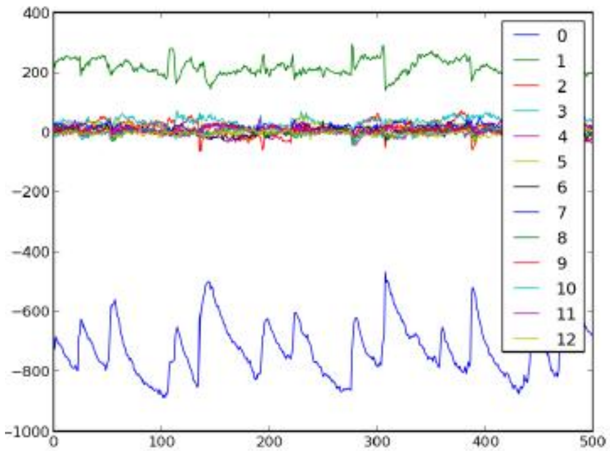
a) parameter 1: Temporal Centroid,
b) parameter 2: Spectral Centroid average value,
c) parameter 3: Spectral Centroid variance,
d) parameters 4-37: Audio Spectrum Envelope (ASE) average values in 34 frequency bands
e) parameter 38: ASE average value (averaged for all frequency bands)
f) parameters 39-72: ASE variance values in 34 frequency bands
g) parameter 73: averaged ASE variance parameters
h) parameters 74,75: Audio Spectrum Centroid – average and variance values
i) parameters 76,77: Audio Spectrum Spread – average and variance values
j) parameters 78-101: Spectral Flatness Measure (SFM) average values for 24 frequency bands
k) parameter 102: SFM average value (averaged for all frequency bands)
l) parameters 103-126: Spectral Flatness Measure (SFM) variance values for 24 frequency bands
m) parameter 127: averaged SFM variance parameters
n) parameters 128-147: 20 first mel cepstral coefficients average values
o) parameters 148-167: the same as 128-147
p) parameters **168-191**: dedicated parameters in time domain based of the analysis of the distribution of the envelope in relation to the rms value.

ISMIS 2011 contest Best Results

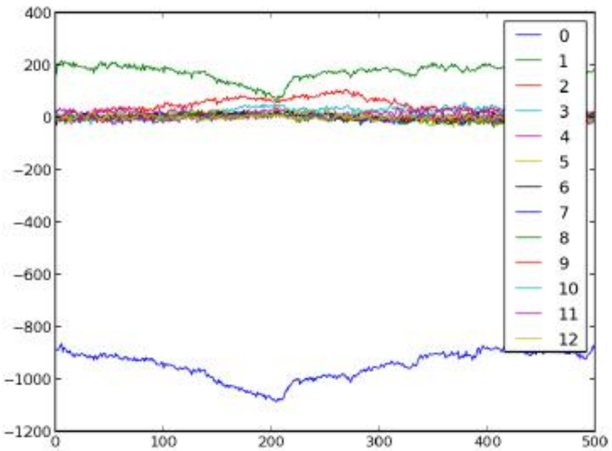
- Amanda Schierz, Marcin Budka and Edward Apeh from Bournemouth University, UK, 1st and 2nd in *Music Genres* track of ISMIS 2011
- *Our classification system consisted of:*
  - *1. Random forest of 1000 unpruned C4.5 decision trees*
  - *2. Boosted ensemble of 10 C5.0 decision trees*
  - *3. Cross-trained ensemble of 100 Naive Bayes classifiers, trained on different subsets of attributes, each time selected using the Floating Forward Feature Selection method.*
- *Result: 0.87270 on the final test set.*

- http://blog.tunedit.org/2011/04/12/domcastro-multiresolution-clustering/

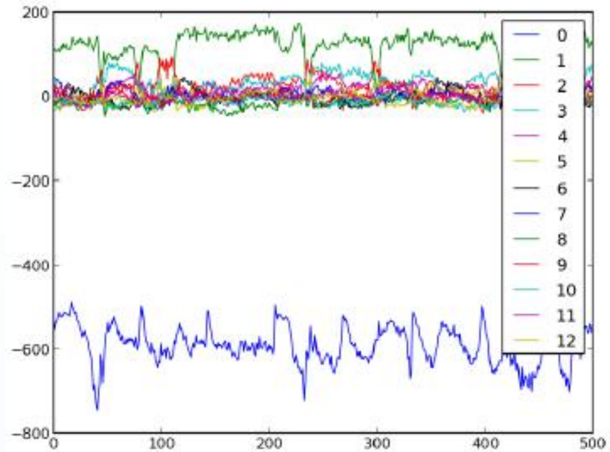
Attributes expressing variations of cepstral coefficients in time



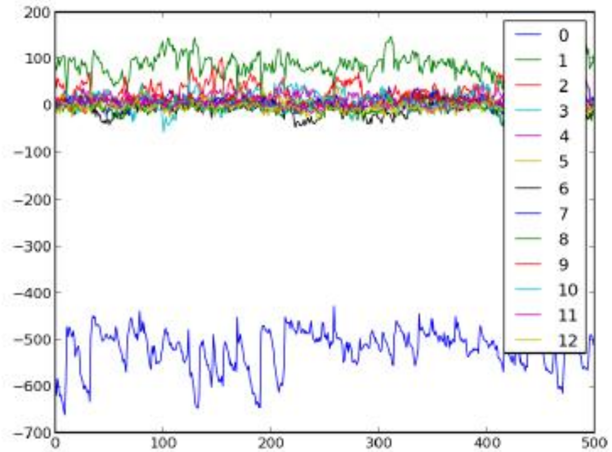
(a) blues



(b) classical



(c) country



(d) disco

Results (1)

- GTZAN collection (1000 files/10 categories)
  - Cross validation results (bagging + Random Forest)

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,940	0,020	0,839	0,940	0,887	0,875	0,996	0,963	classical
0,670	0,017	0,817	0,670	0,736	0,714	0,953	0,790	hiphop
0,730	0,030	0,730	0,730	0,730	0,700	0,951	0,802	reggae
0,760	0,027	0,760	0,760	0,760	0,733	0,953	0,817	country
0,810	0,024	0,786	0,810	0,798	0,775	0,982	0,897	jazz
0,930	0,029	0,782	0,930	0,849	0,835	0,987	0,924	metal
0,730	0,036	0,695	0,730	0,712	0,680	0,946	0,776	disco
0,900	0,026	0,796	0,900	0,845	0,829	0,986	0,917	pop
0,470	0,031	0,627	0,470	0,537	0,500	0,900	0,561	rock
0,680	0,026	0,747	0,680	0,712	0,683	0,964	0,792	blues
0,762	0,026	0,758	0,762	0,757	0,732	0,962	0,824	

0,838

P. Rychły, D. Wiśniewski, E. Łukasik – unpublished results

Results (2)

- http://www-ai.cs.uni-dortmund.de/audio.html database
  - 1886 files /9 classes / unbalanced
  - alternative 145/ blues 120/ electronic 113/folkcountry 222/ funksoulrnb 47/ jazz 319/ pop 116/ raphiphop 300/ rock 504
  - Cross validation results after resampling (training: Bagging + RandomForest)

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.786	0.02	0.833	0.786	0.809	0.956	alternative
	0.974	0.002	0.983	0.974	0.978	0.988	funksoulrnb
	0.692	0.028	0.741	0.692	0.715	0.926	rock
	0.833	0.026	0.802	0.833	0.817	0.983	blues
	0.733	0.042	0.708	0.733	0.72	0.96	jazz
	0.884	0.01	0.918	0.884	0.9	0.984	electronic
	0.767	0.01	0.895	0.767	0.826	0.954	pop
	0.845	0.039	0.735	0.845	0.786	0.972	folkcountry
	0.926	0.016	0.861	0.926	0.892	0.992	raphiphop
Weighted Avg.	0.828	0.022	0.831	0.828	0.828	0.969	

P. Rychły, D. Wiśniewski, E. Łukasik – unpublished results

B.L. Sturm's Survey of Evaluation in Music Genre Recognition

[1] B.L. Sturm, A Survey of Evaluation in Music Genre Recognition, Adaptive Multimedia Retrieval; 01/2012

- Analysis of results of 417 papers related to musical genre recognition

[2] B.L. Sturm and P. Noorzad, On Automatic Music Genre Recognition by Sparse Representation Classification using Auditory Temporal Modulations, Int. Symposium on Computer Music Modeling and Retrieval, 2012

- Critical reproduction of results of [31] Panagakis, Y., Kotropoulos, C., Arce, G.R.: Music genre classification via sparse representations of auditory temporal modulations. In: Proc. European Signal Process. Conf. pp. 1–5. Glasgow, Scotland (Aug 2009)
  presented in the next slide
- „In this work, we review the approach proposed in [31], and describe our attempt to reproduce the results, making explicit the many decisions we have In this work, we review the approach proposed in [31"

B.L. Sturm's results

(a) 5 classes, 88.2% accuracy

Predicted Labels \ True Labels	blues	classical	country	disco	hiphop
blues	0.84	0.00	0.03	0.02	0.00
classical	0.04	0.98	0.08	0.02	0.00
country	0.04	0.02	0.81	0.02	0.00
disco	0.04	0.00	0.06	0.84	0.06
hiphop	0.04	0.00	0.02	0.10	0.94

(b) 10 classes, 66.3% accuracy

Predicted Labels \ True Labels	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock
blues	0.76	0.02	0.07	0.02	0.01	0.07	0.00	0.04	0.05	0.04
classical	0.05	0.89	0.06	0.01	0.01	0.08	0.00	0.02	0.00	0.05
country	0.04	0.01	0.56	0.02	0.01	0.03	0.00	0.02	0.02	0.15
disco	0.02	0.00	0.03	0.45	0.02	0.01	0.00	0.11	0.07	0.04
hiphop	0.03	0.00	0.04	0.15	0.81	0.00	0.02	0.11	0.19	0.04
jazz	0.04	0.02	0.07	0.00	0.01	0.74	0.00	0.03	0.02	0.03
metal	0.00	0.02	0.03	0.07	0.06	0.01	0.93	0.04	0.02	0.14
pop	0.02	0.01	0.02	0.10	0.03	0.02	0.01	0.56	0.07	0.04
reggae	0.02	0.01	0.03	0.08	0.04	0.02	0.00	0.03	0.50	0.04
rock	0.02	0.02	0.09	0.10	0.00	0.02	0.04	0.04	0.06	0.43

Fig. 7. Confusion matrices (frequency of classification) found by stratified 10-fold cross-validation using features derived from the Lyon Ear Model [20, 37] reduced in dimensionality a factor of 4 by PCA, and classification using the LASSO (23).

B.L. Sturm and P. Noorzad, On Automatic Music Genre Recognition by Sparse Representation Classification using Auditory Temporal Modulations, Int. Symposium on Computer Music Modeling and Retrieval, 2012

Violin research
long-standing goal

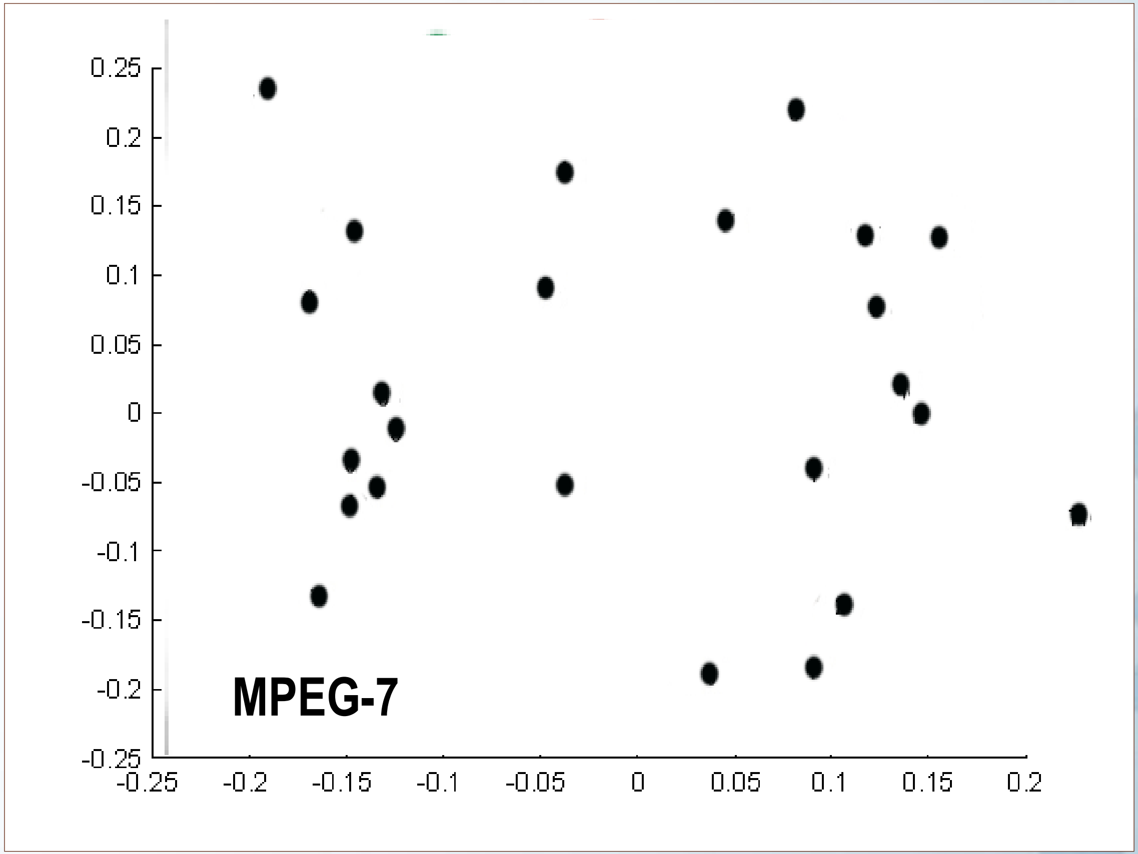
:

To find objective measured attributes
of instruments that correlate
with player-, listener- assessments
of their playing qualities

Motivation

- International Henryk Wieniawski Violinmaker Competition every 5 years in Poznań
- Access to recordings and to jury ranking

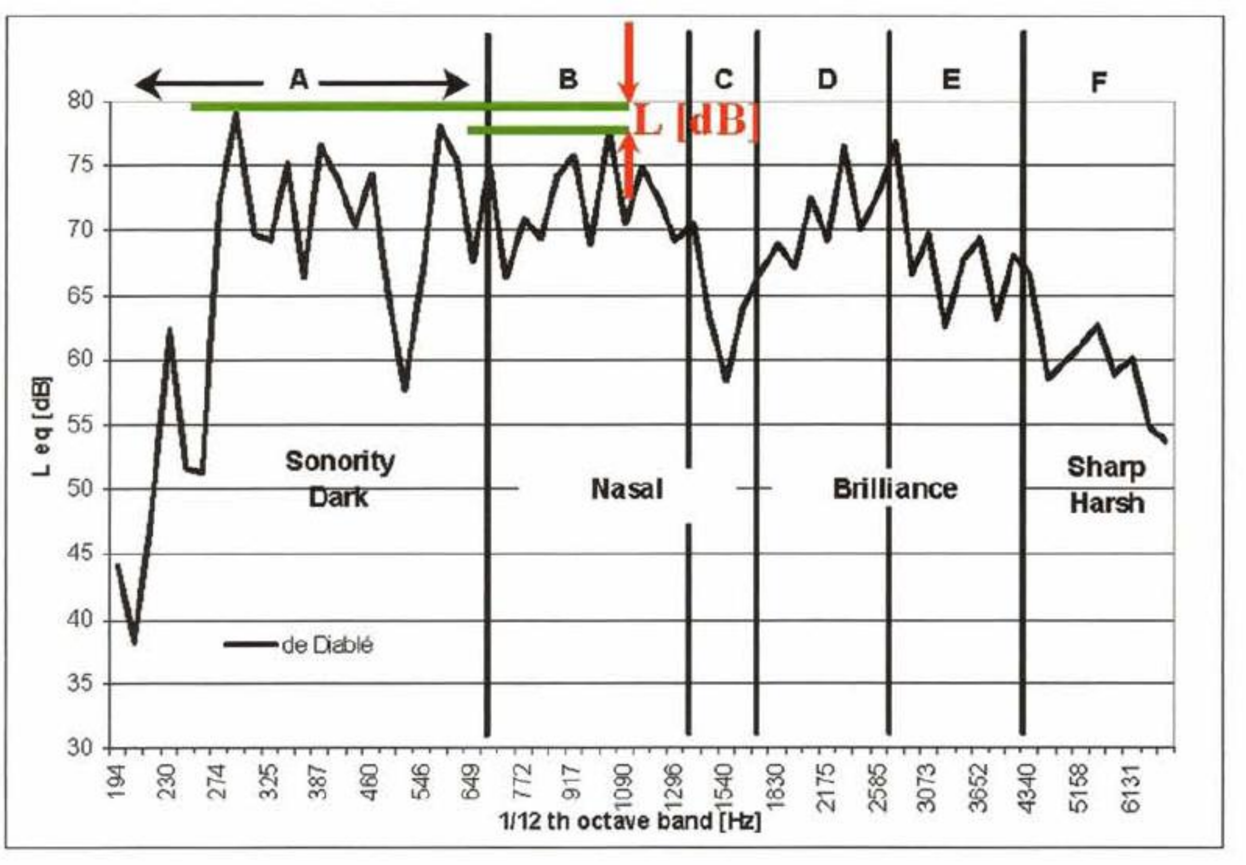
Acoustic features based violin voices diversification



Experts view on the violin sound

- Duennwald, H. (1991). *"Deduction of objective quality parameters on old and new violins,* Catgut Acoust. Soc. J. 1(series II), 1–5
- Fritz C., Blackwell A.F., Cross I., Woodhouse J. & Moore B.C.J. (2012) *„Exploring violin sound quality: Investigating English timbre descriptors and correlating resynthesized acoustical modifications with perceptual properties"*, JASA 131(1), 783-794.
- Buen A., *„What is Old Italian Timbre"*, Proc. of the 2nd Vienna Talk, Univ. of Music and Performing Arts, Vienna, Austria , Sept. 19-21, 2010,
- Zwicker E., *„Psychoacoustics"*, Springer, 1982.
- Assessment of the jury of Henryk Wieniawski Violin competition

Duennwald spectral bands



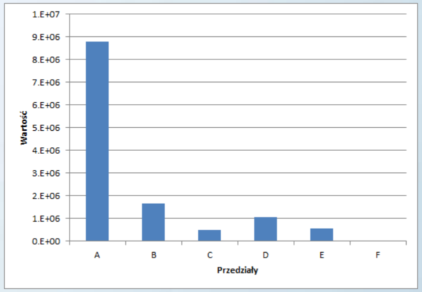
LTAS of the *Violon du Diable, Giuseppe Guarneri del Gesu, 1734*

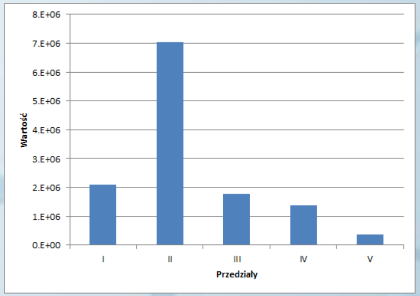
Buen A., On Timbre Parameters and Sound Levels of Recorded Old Violins

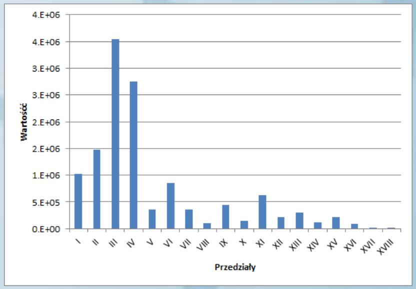
J. Violin Soc. Am.: VSA Papers • Summer 2007 • Vol. XXI, No. 1

Frequency scales used for violins grouping

Duennwald scale	Fritz scale	Bark scale
D1 <190, 650)	**F1** <190, 380)	**B1** <190, 300)
		B2 <300, 400)
	F2 <380, 760)	**B3** <400, 510)
		B4 <510, 630)
D2 <650, 1300)		**B5** <630, 770)
	F3 <760, 1520)	**B6** <770, 920)
		B7 <920, 1080)
		B8 <1080, 1270)
D3 <1300, 1640)		**B9** <1270, 1480)
	F4 <1520, 3040)	**B10** <1480, 1720)
D4 <1640, 2580)		**B11** <1720, 2000)
		B12 <2000, 2320)
D5 <2580, 4200)		**B13** <2320, 2700)
		B14 <2700, 3150)
	F5 <3040, 6080)	**B15** <3150, 3700)
		B16 <3700, 4400)
D6 <4200, 7000)		**B17** <4400, 5300)
		B18 <5300, 6400)







Next steps: finding similarity, decision trees for grouping, GMM modeling…

Other research assumptions

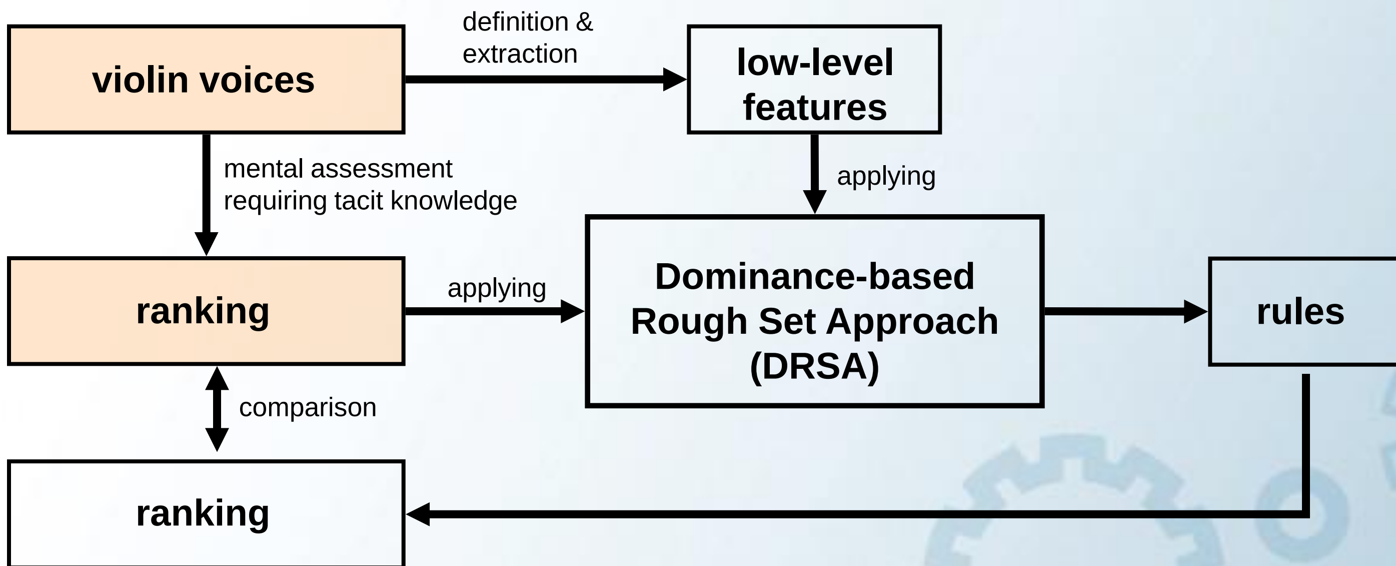
Questions:

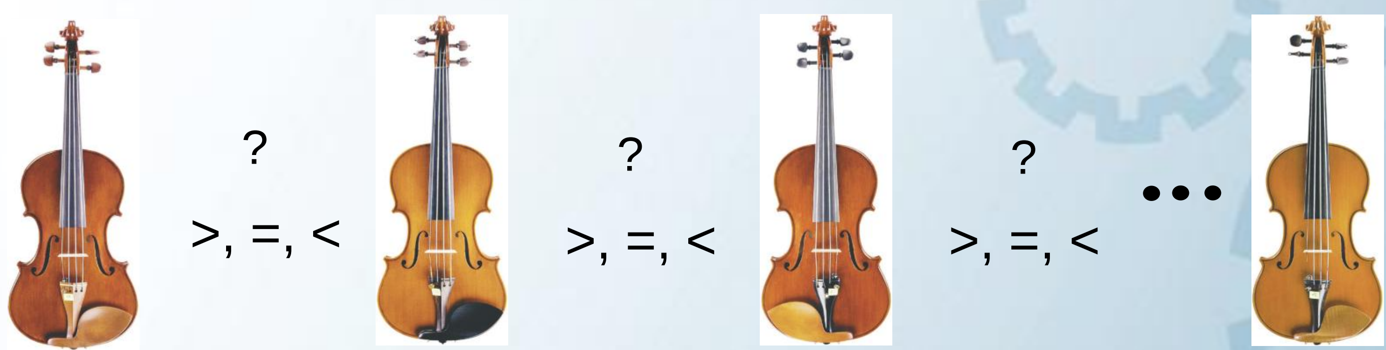
- Is it possible to acquire experts' tacit knowledge which is subconsciously used in expression of their preferences?
- What features of the violin sound are relevant to the mental process of making a ranking?

Our aim is to discover an **expert's preference model** which is usually a part of his/her tacit knowledge. The preference model should:

- explain an expert's ranking policy,
- be inferred from examples of expert's judgments,
- be expressed in a formal way, e.g. by a set of rules based on the objects' features,
- allow to rank new objects with high accuracy (good generalization required).

Genaral goal

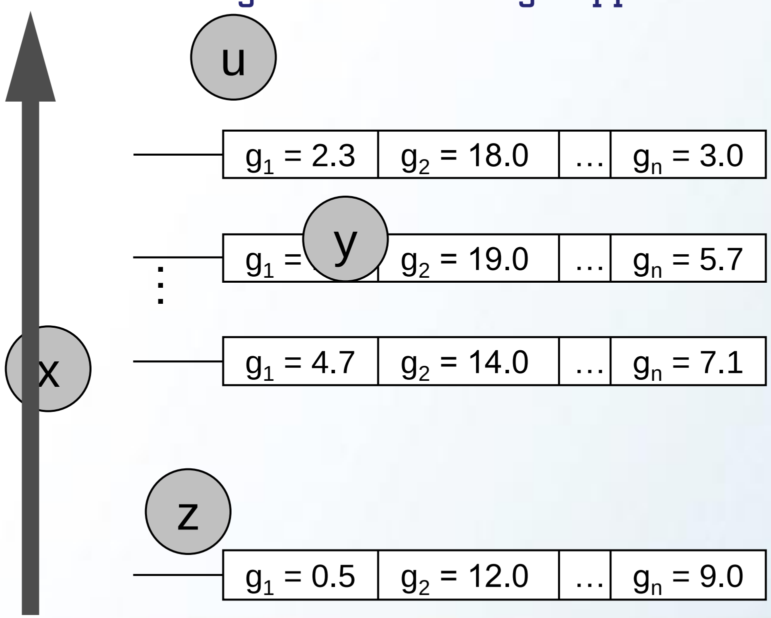




Greco, S., Matarazzo, B., Słowiński, R.: Multicriteria classification by dominance-based rough set approach. In: W.Kloesgen and J.Zytkow (eds.), Handbook of Data Mining and Knowledge Discovery, Oxford University Press, New York, 2002

Methodology of acquiring the preference model based on Dominance-based Rough Set Approach

- acquiring multimedia objects (pictures, sound files, tablet data…)
- estimation of the objects:
  - making a ranking
  - pairwise comparison of the objects, i.e. for a pair (x, y) the decision maker is asked to specify whether object x is, comprehensively, at least as good as object y (x **S** y) or not (x **S**^{c} y)
- providing and computing the objects' features,
- creating the Pairwise Comparison Table (PCT),
- inducing rules from rough approximations of relations **S** and **S**^{c}

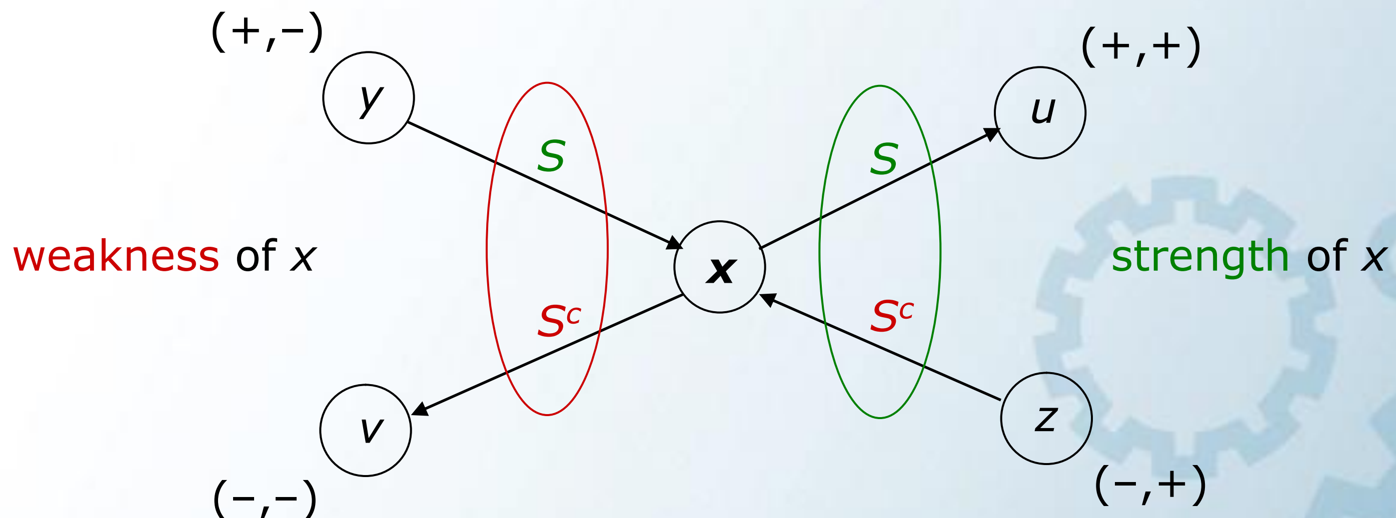


Pairs of objects	Difference in the values of features				Preference relation
	Δ **g$_1$**	Δ **g$_2$**	…	Δ **g$_n$**	
(x,y)	$\Delta_1(x,y) = -1.1$	$\Delta_2(x,y) = 1.0$	…	$\Delta_n(x,y) = 2.7$	x **S^c** y
(y,z)	$\Delta_1(y,z) = -2.4$	$\Delta_2(y,z) = 4.0$	…	$\Delta_n(y,z) = -4.1$	y **S** z
(y,u)	$\Delta_1(y,u) = 1.8$	$\Delta_2(y,u) = 6.0$	…	$\Delta_n(y,u) = -6.0$	y **S** u
…	…	…	…	…	…
(u,z)	$\Delta_1(u,z) = -4.2$	$\Delta_2(u,z) = -2.0$	…	$\Delta_n(u,z) = 1.9$	u **S^c** z

Multicriteria ranking – Net Flow Score

S – positive argument

S^c – negative argument



$NFS(x)$ = strength(x) – weakness(x)

ranking: complete preorder determined by $NSF(x)$

Practical verification

jury's assessment

sound recording

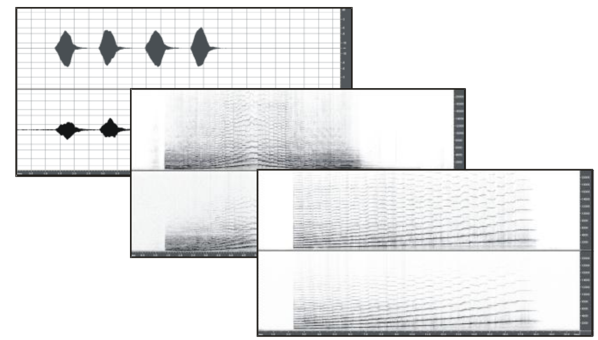
criteria:
- **volume of sound (X)**,
- **timbre of sound (Y)**,
- easy of sound emission,
- **equal sound volume of strings (Z)**,
- accuracy of assembly,
- individual qualities

Ranking of violins based on the criterion **X**

Ranking of violins based on the criterion **Y**

Ranking of violins based on the criterion **Z**

The violin's acoustic data:
- individual sounds played on open strings, G,D,A,E,
- successive sounds of chromatic scale,



Acoustic features:
- power spectrum of chromatic scale sounds,
- wavelets,
- harmonic based spectral parameters (tristimuli, brightness, odd/even harmonics content...),
- psychoacoustic features
- cepstral coefficients.

Dominance-based Rough Set Approach

Exemplary DRSA results

- reconstructing the expert's rankings of set of 23 violins,
- three rankings: volume, timbre and inter-string equality,
- feature space - cepstral coefficients for each string (G, D, A, E)

criterion	best subset of features (cepstral coefficients)	number of rules	ranking distance
volume	A14, E13, D12, G16	62	1.57
timbre	E13, D15, G4, G17, D5	99	2.00
inter-string equality	D20, D15, A24, D10	64	2.26

Jelonek, J., Łukasik, E., Naganowski, A., Słowiński, R, Inferring Decision Rules from Jury's Ranking of Competing Violins, Proc. SMAC'03, Stockholm, 2003, 75-78

Exemplary DRSA results (2)

- reconstructing the expert's rankings of set of 23 violins,
- three rankings: volume, timbre and inter-string equality,
- feature space – equal loudness curves composed of the energy of individual sounds of the chromatic scale played on each string

criterion	best subset of features Sounds of the chromatic scale played on each string	number of rules	ranking distance
volume	G:G#4, G:D5, A:F#6, A:G6, E:A#5	98	2.08
timbre	D:G4, A:D#6, A:G#6, E:F6, E:D# 7	82	1.87
inter-string equality	G:G#4, D:F5, D:A5, A:D:6	57	1.70

Jelonek, J., Łukasik, E., Naganowski, A., Słowiński, R, Inducing jury's preferences in terms of acoustic features of violin sounds, Lecture Notes in Artificial Intelligence, Springer 2004, 492-497

The software tool „Ranker"

„Ranker" has been created to enable:

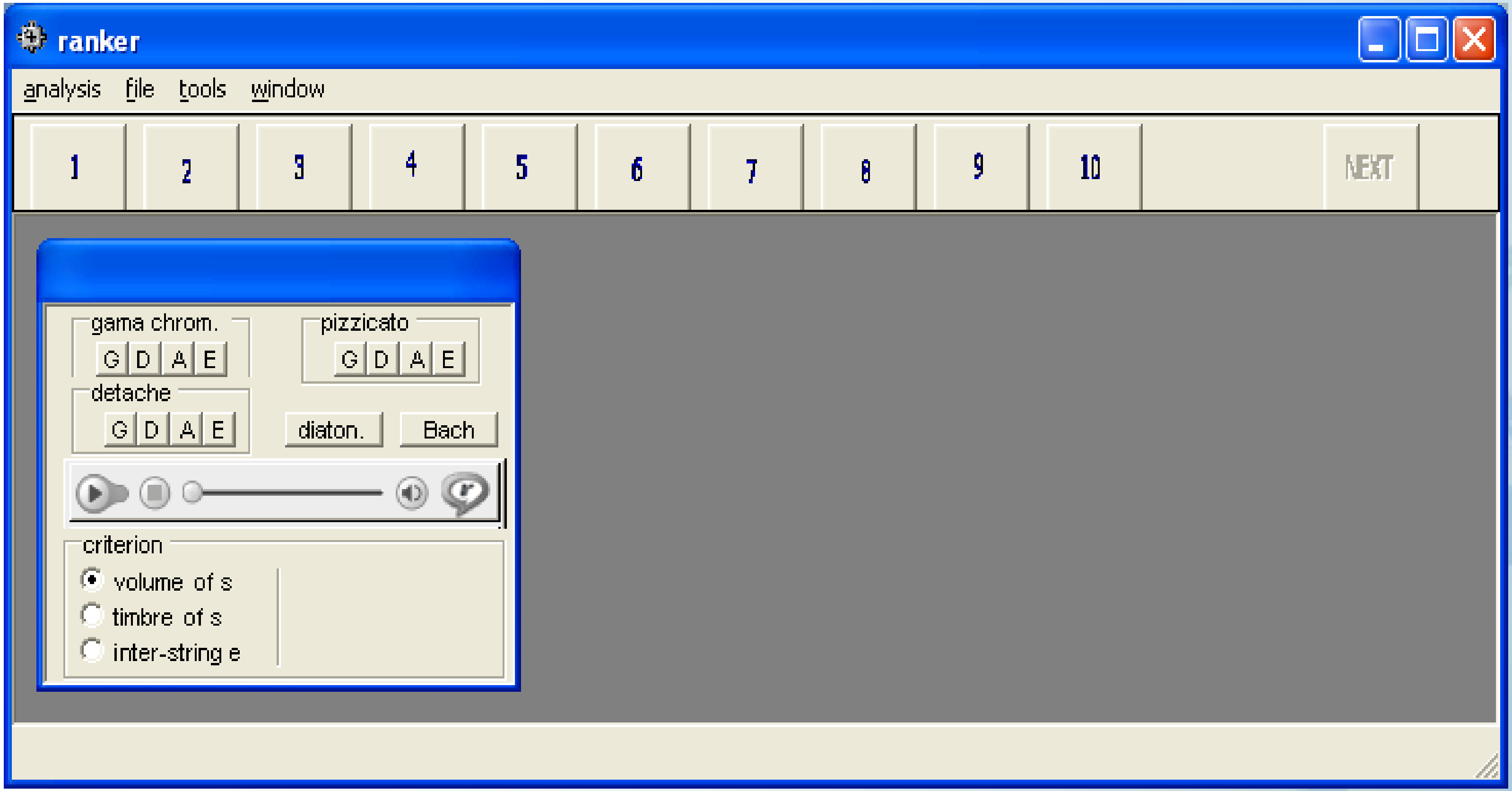
- the presentation of stimuli to the expert,
- the creation of ranking of audio (and visual) objects,
- comparison of ranking method used,
- the choice of:
  - data sets,
  - method of assessment (single-stimulus, multistimulus,
  - pairwise comparison),
  - number and type of criteria used,
  - number of scale levels for scoring.

Budzyńska L., Jelonek J., Łukasik E., Susmaga R., Słowiński R., Multistimulus ranking versus pairwise comparison in assessing quality of musical instruments sounds, AES Convention Paper no 137, Barcelona, 2005.

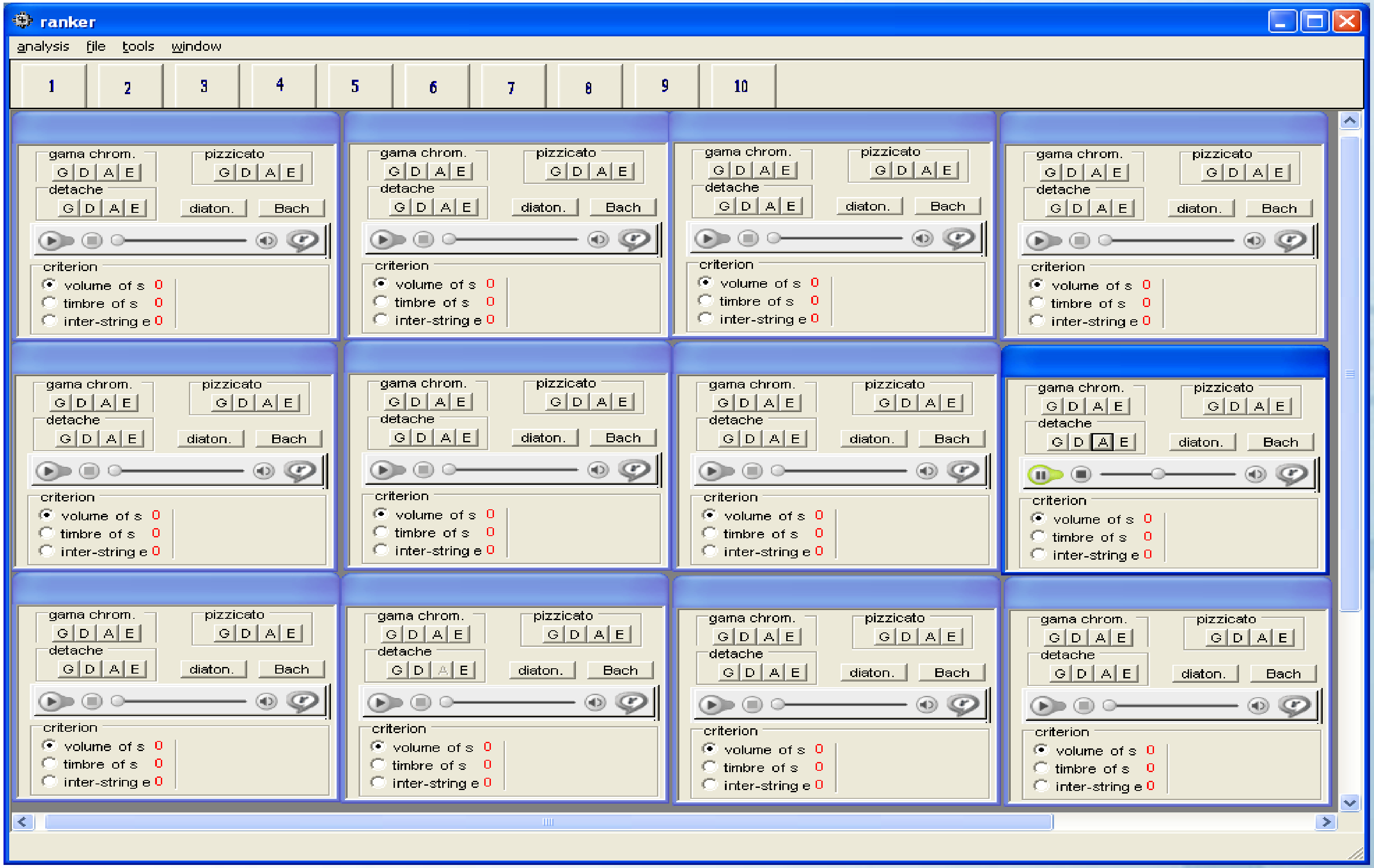
Ranking methods

- Single stimulus ranking
  - allows the expert to assess individual objects presented to her/him in random order by the system,
- Pairwise comparison
  - the expert is confronted with two objects to be evaluated (compared) at a time ($<, \ = , >$)
- Multistimulus ranking
  - inspired by the MUSHRA method (Multistimulus with Hidden Reference and Anchor)

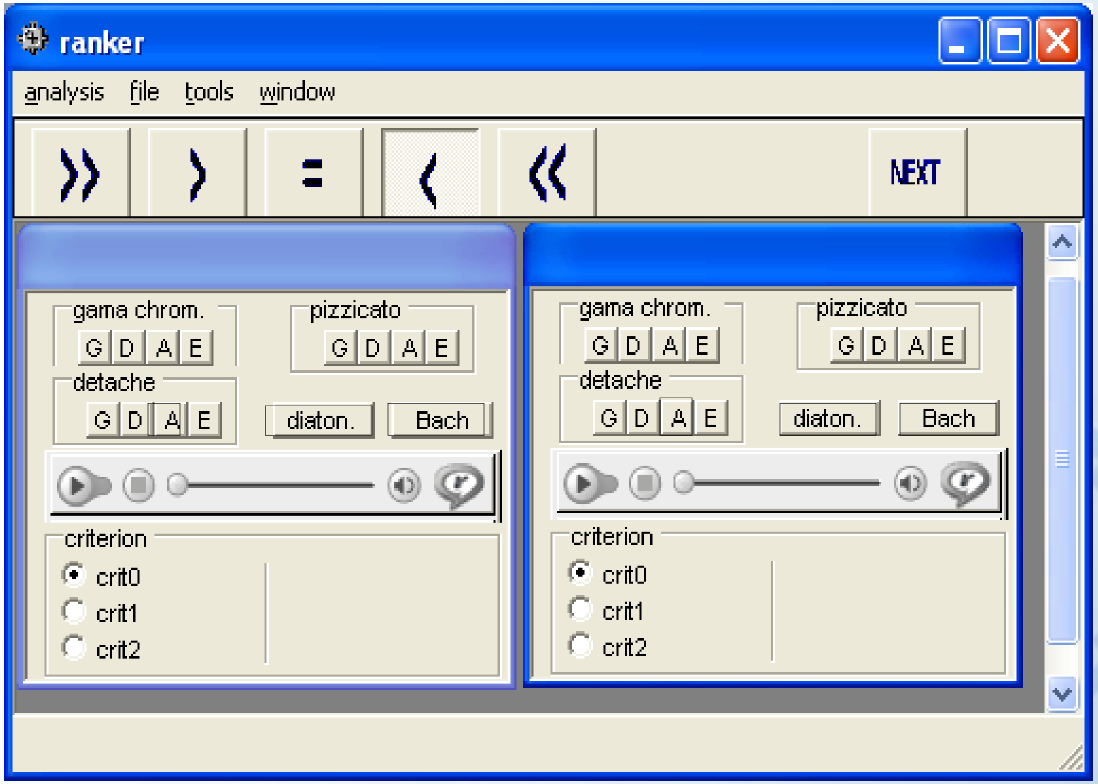
„Ranker” - single-stimulus ranking



„Ranker" - multistimulus ranking



„Ranker" – pairwise comparison



Pairwise Comparison Table (PCT) followed by Net Flow scoring (NFS) are implemented in „Ranker".

PCT and NFS in „Ranker"

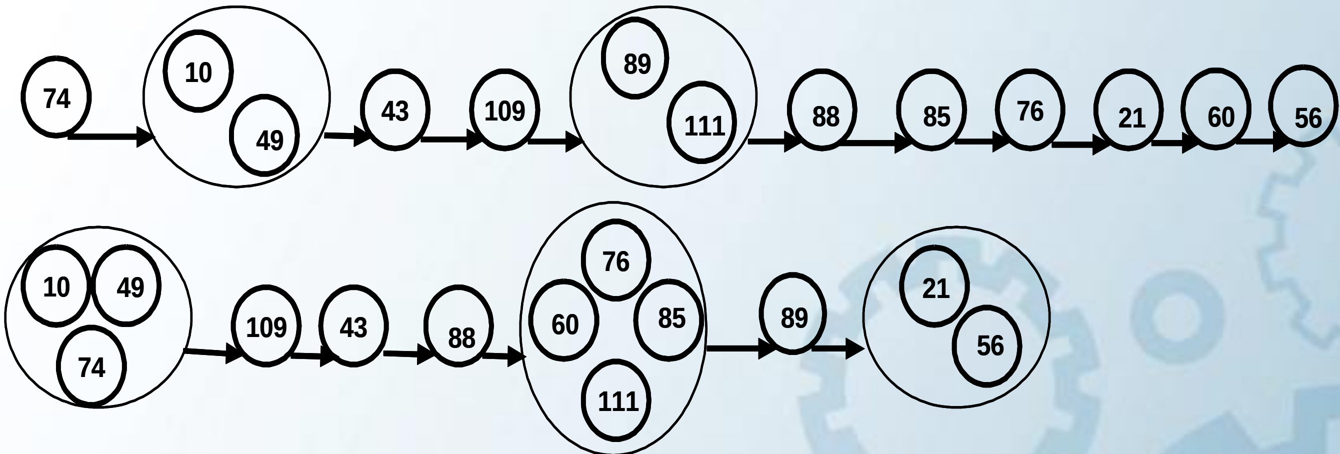
ranking

alternatives	sila	barwa	wyrown...
60	-9	-9	2
56	-9	0	-11
76	-7	2	-1
85	-7	5	0
89	-5	12	6
111	0	1	-1
21	1	-9	-4
109	2	-6	1
88	4	-7	-2
74	5	2	2
10	7	-2	5
49	7	6	4
43	11	5	-1

Results - ranking

Pairwise comparison

Multistimulus ranking



Expert CC

Results - similarity of rankings

Kendall coefficient - τ

Criterion\Expert	BB	CC
volume	0.54	0.65
timbre	0.38	0.59
inter-string	0.22	0.35
vol.+tim.	0.47	**0.69**
vol.+tim.+inter	0.38	0.65

Blest coefficient - ν

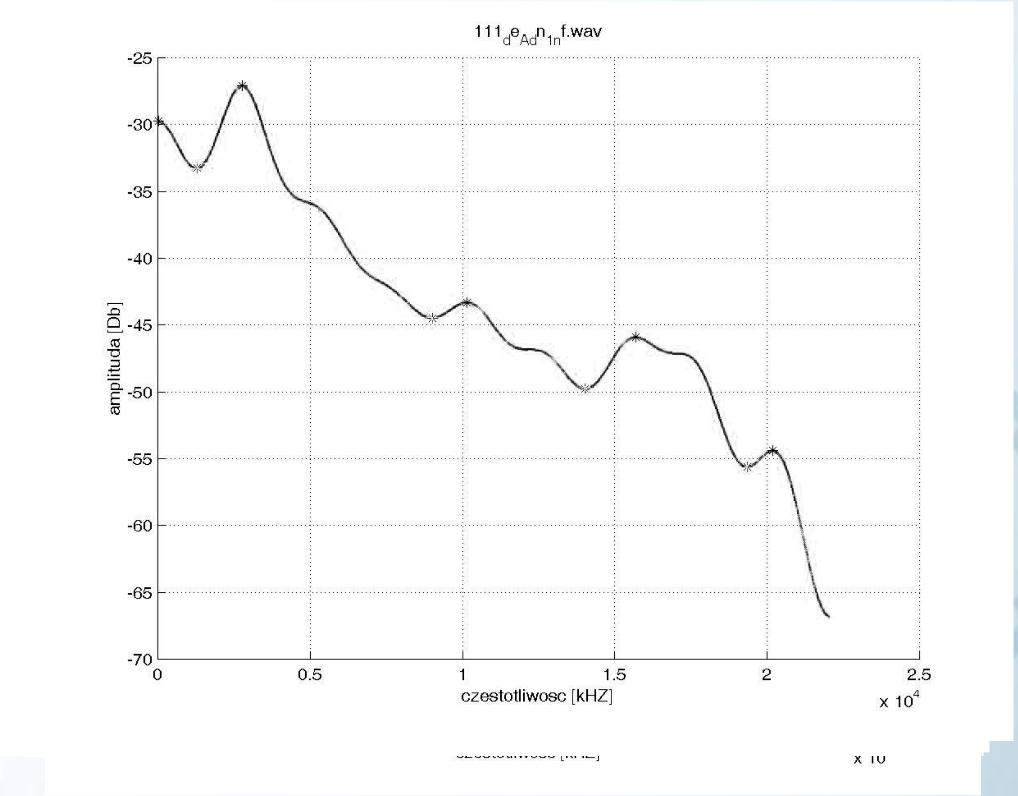
Criterion\Expert	BB	CC
volume	**0.80**	**0.81**
timbre	**0.68**	0.64
inter-string	0.46	0.21
vol.+tim.	0.62	**0.79**
vol.+tim.+inter	0.55	0.74

Bibliography on DRSA and its use for violin

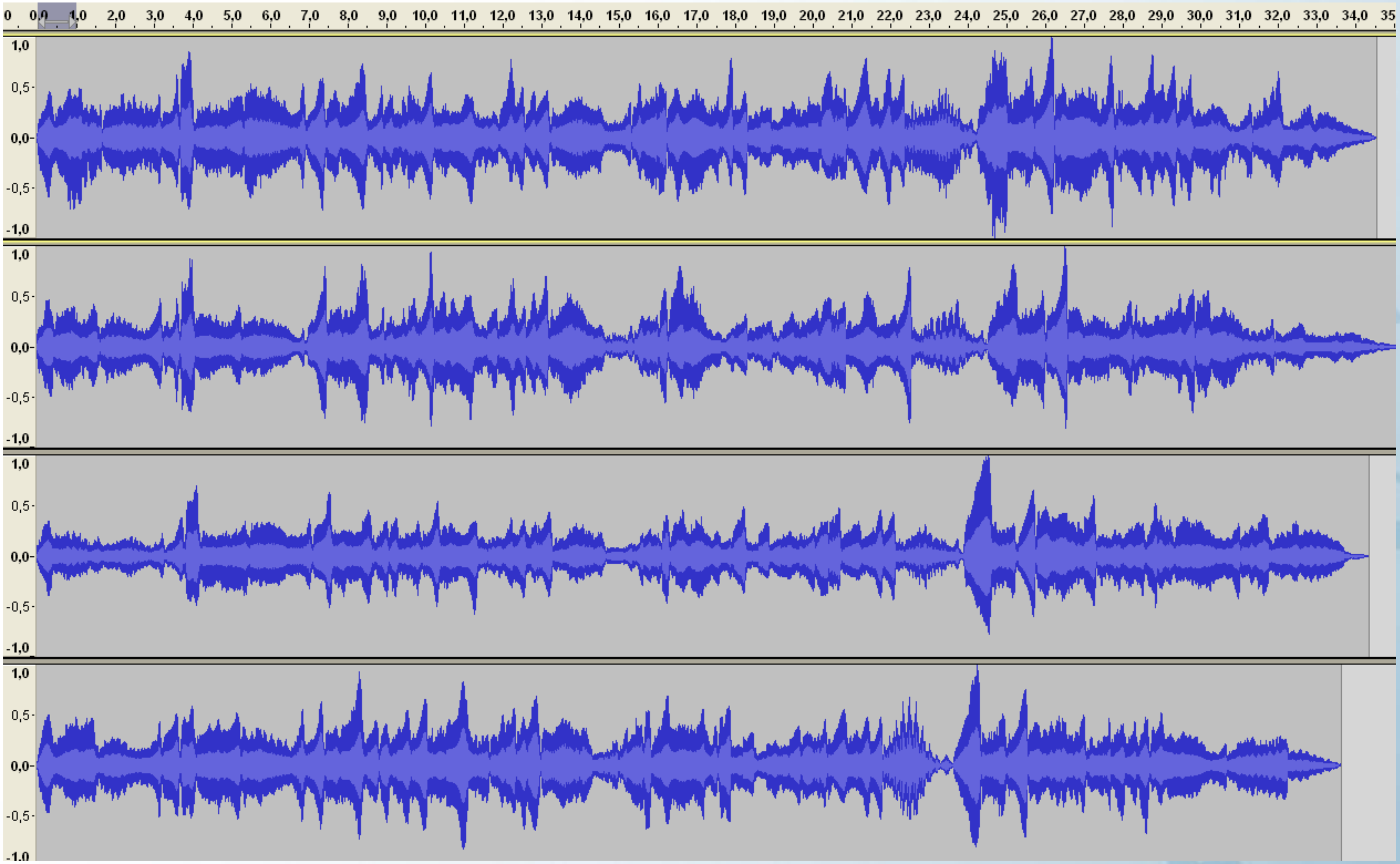
- Greco, S., Matarazzo, B., Słowiński, R.: Multicriteria classification by dominance-based rough set approach. In: W.Kloesgen and J.Zytkow (eds.), Handbook of Data Mining and Knowledge Discovery, Oxford University Press, New York, 2002
- Słowiński R., Greco, S., Matarazzo, B.: Rough set based decision aiding. [In]: E. Burke and G. Kendall (eds.), Introductory Tutorials on Optimization, Search and Decision Support Methodologies, Kluwer Academic Publishers, Boston, 2003, chapter 16
- Słowiński, R., Greco, S., Matarazzo, B.: Mining decision-rule preference model from rough approximation of preference relation. [In]: Proc. 26th IEEE Annual International Conference on Computer Software & Applications, 26-29 August 2002, Oxford, England, 1129-1134
- Jelonek, J., Łukasik, E., Naganowski, A., Słowiński, R, Inferring Decision Rules from Jury's Ranking of Competing Violins, Proc. SMAC'03, Stockholm, 2003, 75-78
- -------------------Inducing jury's preferences in terms of acoustic features of violin sounds, Lecture Notes in Artificial Intelligence, Springer 2004, 492-497
- Budzyńska L., Jelonek J., Łukasik E., Susmaga R., Słowiński R., Multistimulus ranking versus pairwise comparison in assessing quality of musical instruments sounds, AES Convention Paper no 137, Barcelona, 2005.
- ----------Knowledge Discovery from Digital Media Objects Using Preference Semantics, in: Knowledge-Based Media Analysis for Self-Adaptive and Agile Multimedia Technology, Proc. EWIMT 2004, Queen Mary Univ. London 2004, 15-23

Instead of conclusions

invitation to find similarity between cepstral smoothing curves . . .



... and similarities of some other shapes (waveforms) ...



Thank you
for your attention

????