Multimedia Systems

IV. Music Information Retrieval

4.1. Timbre

António Sá Pinto

FEUP – 2019/2020

Agenda

- Timbre - Summary
  - How to analyse it?
  - How to describe it?
  - Main descriptors
- Applications

Timbre

Definition

The attribute of sensation in terms of which a listener can judge that two sounds have the same loudness and pitch are dissimilar. *(ANSI)*

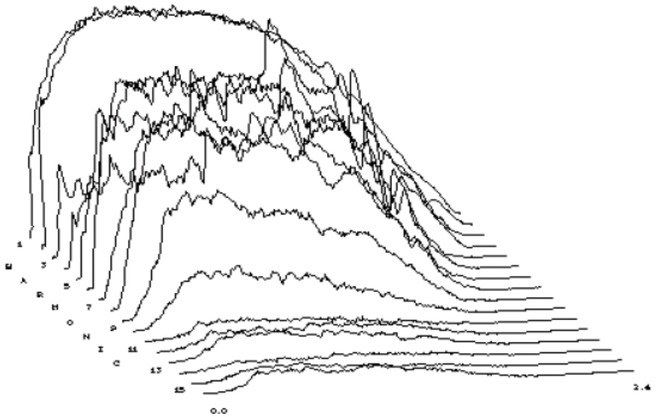
- In other words, everything that is not loudness or pitch! *...the psychoacoustician's multidimensional waste-basket category for everything that cannot be labeled pitch or loudness.*" (McAdams and Bregman 1979)

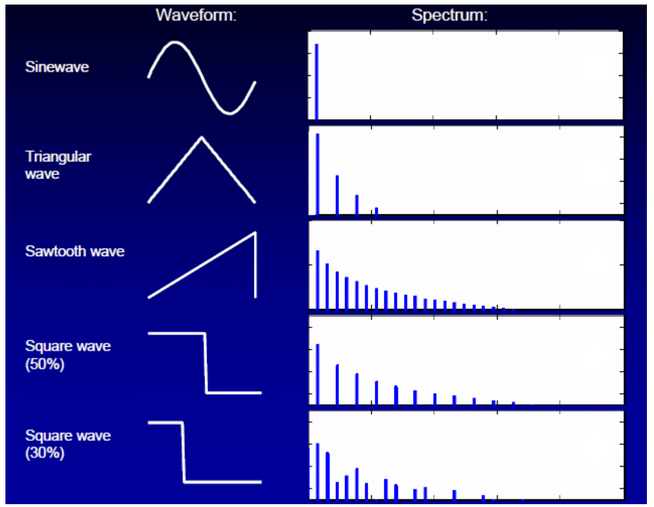
- Musically, it is essential to distinguish different types of musical instruments, e.g. to distinguish a *violin* playing a G2 (98Hz) note and a *trumpet* playing the same note (98Hz), at equal loudness and duration.

Timbre

Definition

relates to: static spectrum; spectral envelope; time envelope (ADSR); dynamic spectrum (time-evolving); phase; etc,…





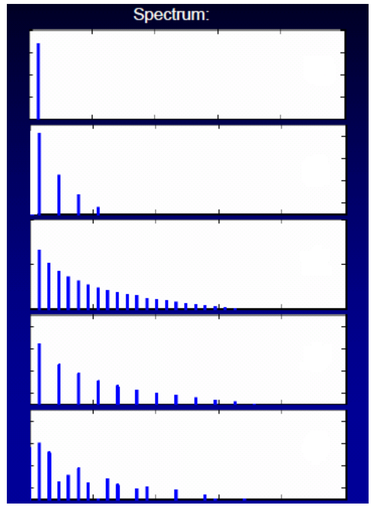
- Not ”measured”, given its perceptual nature;
- Not even ”estimated”, given its multidimensional nature;
- Can be “described” computationally – descriptors.

Timbre

How to analyse it?

relates to:

- static spectrum;
- spectral envelope;
- dynamic spectrum (time-evolving);
- time envelope (ADSR model);
- phase;
- Etc;

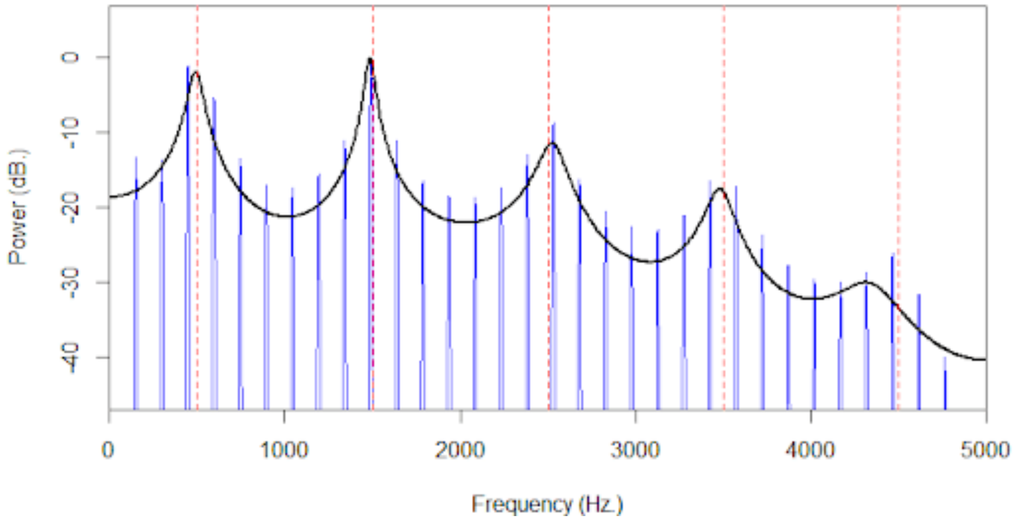


Timbre

How to analyse it?

relates to:

- static spectrum;
- **spectral envelope;**
- dynamic spectrum (time-evolving);
- time envelope (ADSR model);
- phase;
- Etc;



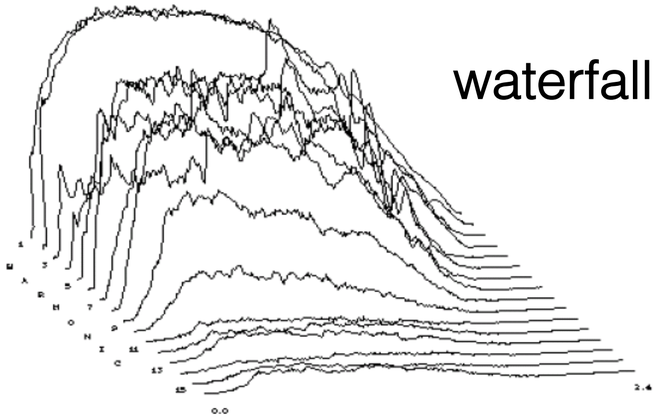
Timbre

How to analyse it?

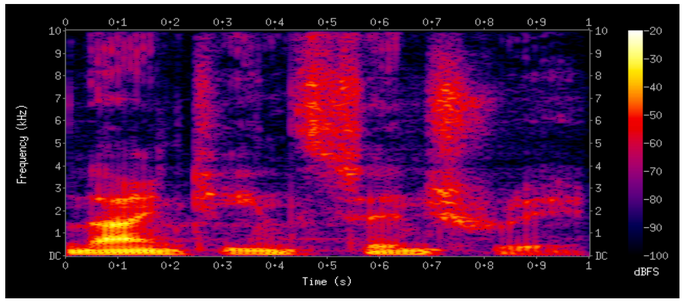
relates to:

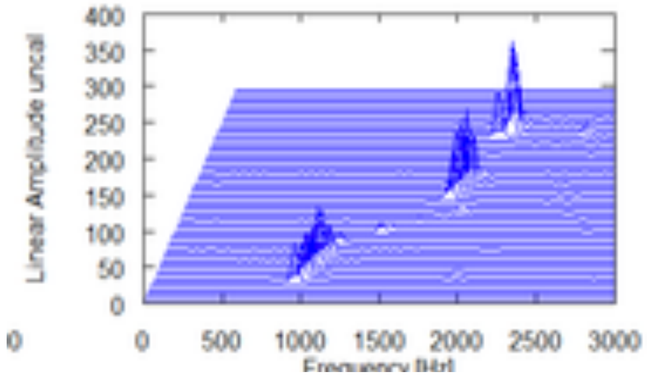
- static spectrum;
- spectral envelope;
- dynamic spectrum (time-evolving);
- time envelope (ADSR model);
- phase;
- Etc;

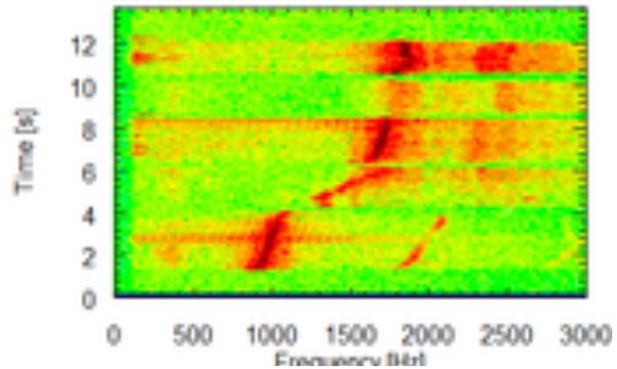
waterfall



spectrogram







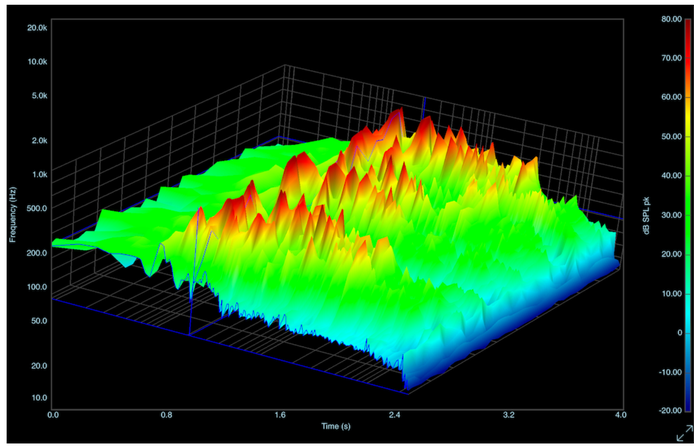
Timbre

How to analyse it?

relates to:

- static spectrum;
- spectral envelope;
- dynamic spectrum (time-evolving);
- time envelope (ADSR model);
- phase;
- Etc;

Mixed waterfall-spectrogram



Timbre

relates to:

- static spectrum;
- spectral envelope;
- dynamic spectrum (time-evolving);
- time envelope (ADSR model);
- phase;
- Etc;

How to analyse it?

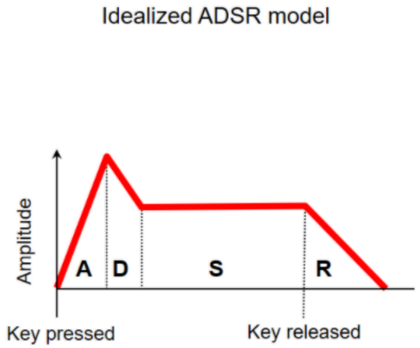
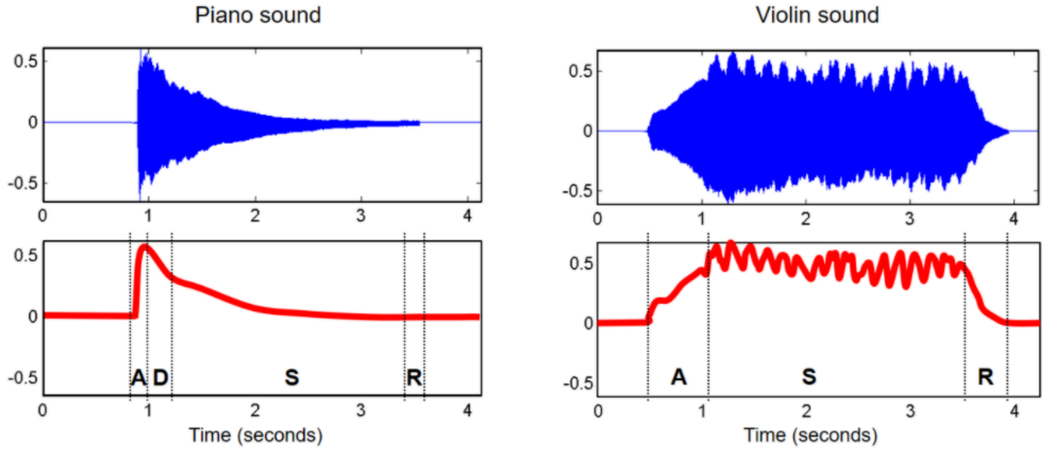


Figure 1.22b and Figure 1.23
from [Müller, FMP, Springer 2015]



Timbre

relates to:
- static spectrum;
- spectral envelope;
- dynamic spectrum (time-evolving);
- **time envelope (ADSR model);**
- phase;
- Etc;

Idealized ADSR model

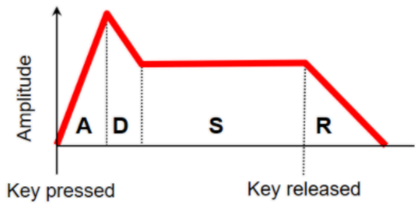


Figure 1.22b and Figure 1.23
from [Müller, FMP, Springer 2015]

How to analyse it?

Why?

- Perceptually very important (namely **attack**)
- Distinct zones have very distinct properties (descriptors)

Timbre

How to analyse it?

relates to:
- static spectrum;
- spectral envelope;
- dynamic spectrum (time-evolving);
- time envelope (ADSR model);
- phase;
- Etc;

QUESTION

What type(s) of instrument may be modelled by a simple **AR** model?
(without Decay and Sustain)

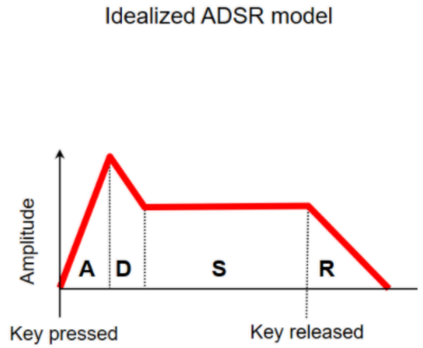
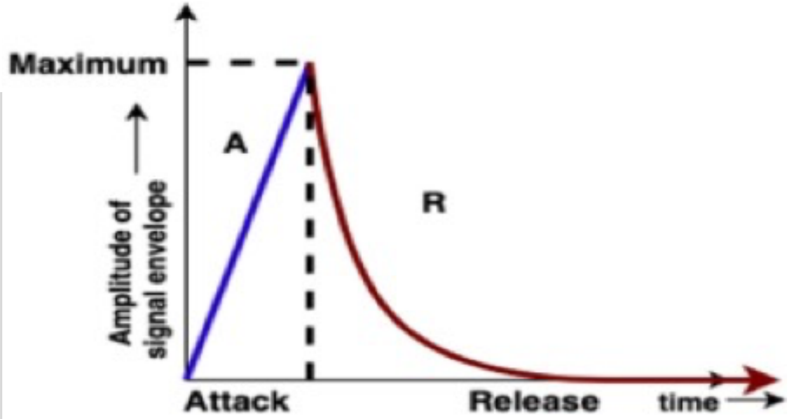


Figure 1.22b and Figure 1.23
from [Müller, FMP, Springer 2015]



Timbre

Main Descriptors

Descriptor	Domain	Scope	Indications	Typical Values
Log-Attack Time	Temporal	Global	Requires modelling (envelope). Only valid when onsets are located	*~15 ms guitar; 20 ms harp; kick 3ms, snare 1ms*
Temporal Centroid	Temporal	Global	Related to decay time	*non-sustained: guitar, harp, pizzicato violin (~20%) vs sustained (~50% of duration)*
Zero-Crossing Rate	Temporal	Instantaneous	Correlated with spectral centroid. Approximate indicator of pitched or unpitched content	low for harmonic; high for percussive; (sensitive to noise)
RMS Energy	Temporal	Instantaneous	LOUDNESS (not TIMBRE)	
FreqMod	Temporal	-	Vibrato (Requires modelling)	
AmpMod	Temporal	-	Tremolo (Requires modelling)	
Spectral Centroid	Spectral	Instantaneous	Indicator of brightness ("centre of gravity of spectrum")	BRIGHTNESS
Spectral Spread (2nd)	Spectral	Instantaneous	Spectral bandwidth	
Spectral Flux	Derivative	Instantaneous	Ammount of spectral change among frames	
MFCC			Wide-spread timbre descriptors (speech recognition)	Lower coefficients represent spectral envelope; Higher ones are related to finer details of spectrum
Spectral Flatness	Spectral	Instantaneous	Measure of noisiness. Computed for different frequency bands	0 for tonal (sinusoid), 1 for noise
Noisiness	Model		Energy of noise over total energy	
Odd-to-even	Model		Odd-to-even harmonic energy ration	
Tristimulus	Model		The audio counterpart for RGB (not very used)	

And many many others…

Timbre

Applications

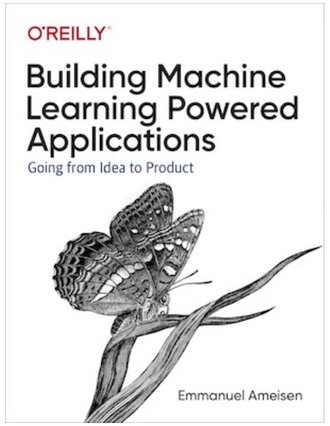
- Instrument Classification;
- Sound Morphing;
- Song Identification;

- Many Others:
  - Timbre Spaces;
  - Sound Visualization;
  - Description of Large Libraries;
  - Automatic Creation of Playlists (e.g. Spotify);
  - Etc.

Timbre

Instrument Classification (I)

- Be the algorithm!



If you have never tried doing your algorithm's job, it will be hard to judge the quality of its results. On the other side, if you spend some time labeling data yourself, you will often notice trends that will make your modeling task much easier.

You might recognize this advice from our previous section about heuristics, and it should not surprise you. Choosing a modeling approach involves making almost as many assumptions about our data as building heuristics, so it makes sense for these assumptions to be data driven.

Timbre

Not covered

Covered

Instrument Classification (II)

Taxonomy → Dataset Creation → Feature Extraction → Classification → Evaluation

Definition

- Instrument Taxonomies;
- Playing method!!
- Pitched vs non-pitched;
- Sustained vs non-sustained;

Annotation

- Instrument, note,
- Playing method!!
- Pitched vs non-pitched;
- Sustained vs non-sustained;

Gather sources

- Representative of the problem
- Sound quality;

Extraction

Transformation

Projection

Selection

Approaches

- Single classifier for all instruments;
- Hierarchical /ensembles of classifiers;

Automatic

- Unsupervised/Supervised
- Algorithms: SVM, K-NN, etc.

Manual

IR Measures

- Precision,
- Recall,
- Accuracy,
- F-Measure

Timbre

Instrument Classification (III)

Author, year [ref]	Total instances[a]	NC	Acoustic features[b]	Classification algorithm	Instrument performance (%)	Family performance (%)
Eronen, 2001 [174]	5286 (MUMS, Iowa, SOL, RolandXP30, own recordings)	29	MFCC (attack-steady), F0, ATT, onset features, SC, Crest Factor, AM	k-NN	35 (H: 30)	77 (H: 75)
Livshin et al., 2003 [413]	4381 (SOL, Iowa, MUMS, Prosonus, Vitous)	16	SC, ATT, temporal decrease, TRI, HD, SKW, KUR, SV, SS, MFCC, noisiness	LDA & k-NN	47-69	62-92
Peeters, 2003 [511]	4163 (SOL, Iowa, MUMS, Microsoft MI, Prosonus, Vitous)	23	same as above	LDA & GMM (hierarchical)	54 (H: 64)	81 (H: 85)
Eronen, 2003 [175]	5895 (MUMS, Iowa, SOL, Martin, own recordings)	7	MFCC, delta-MFCC + ICA	HMM	68	n/a
Kitahara et al., 2003 [345]	6247 (RWC)	19	SC, OER, F0 relative energy, KUR, SKW, FM, amplitude envelope slope, onset energy	Bayes (k-NN after PCA & LDA)	80	91
Kostek et al., 2004 [362]	n/a (CMIS, MUMS)	12	Wavelet-based energy bands, MPEG-7 features	MLP	71	n/a
Szczuko et al., 2004 [615]	2517 (CMIS, MUMS)	16	MPEG-7 features, OER, F0	MLP (2-stage hierarchical MLP)	86 (H: 89)	n/a
Park et al., 2005 [496]	829 (several commercial instrument-sample CDs)	12	SS, SC, harmonic slope, LPC noise, harmonic expansion/contraction, spectral jitter and shimmer, spectral flux, TC, ZCR	MLP with elliptical/ radial basis functions	71	88
Chétry et al., 2005 [88]	4415 (Iowa, RWC, voice)	11	Line spectrum frequencies	K-means derived codebook	95	n/a

Timbre

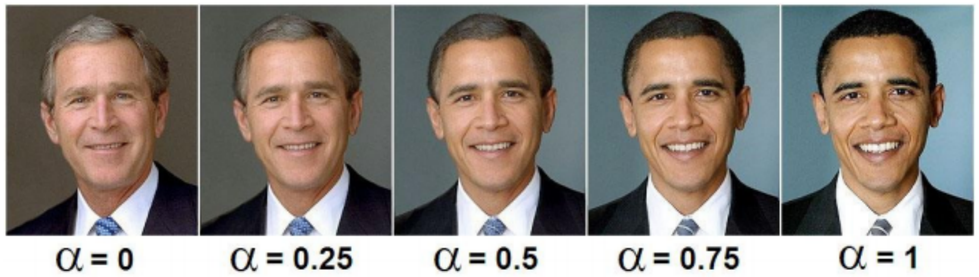
Sound Morphing (I)

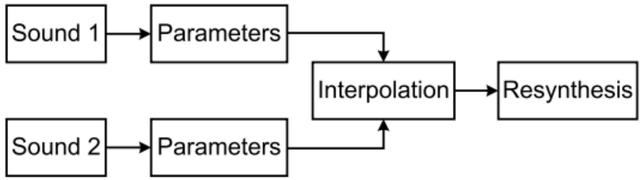
- transformation
- smoothly transition from one sound to another across timbre dimensions
- Mixing is not morphing

Timbre

Sound Morphing (II)

- smoothly transition from one sound to another
- Mixing is not morphing





$$S_{12} = \alpha S_1 + [1 - \alpha]\, S_2$$

from: (Caetano and Rodet) Automatic Timbral Morphing of Musical instrument Sounds by High-Level Descriptors. *ICMC Conference, 2010*

Timbre

Sound Morphing (III)

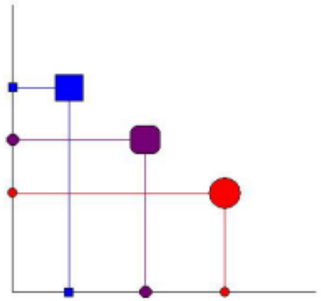
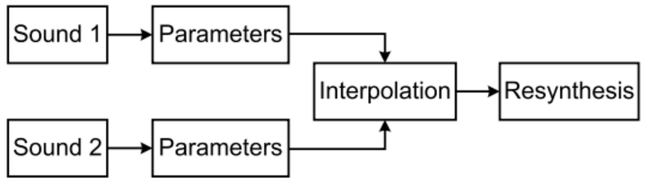
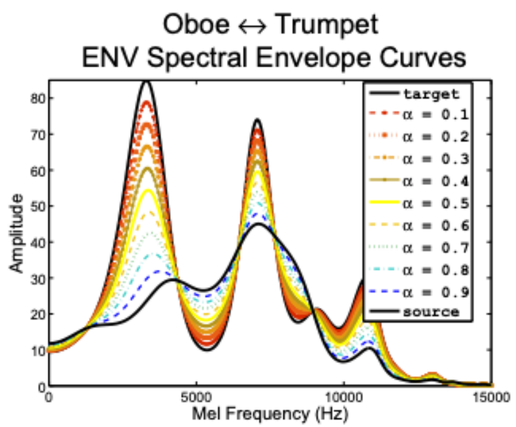
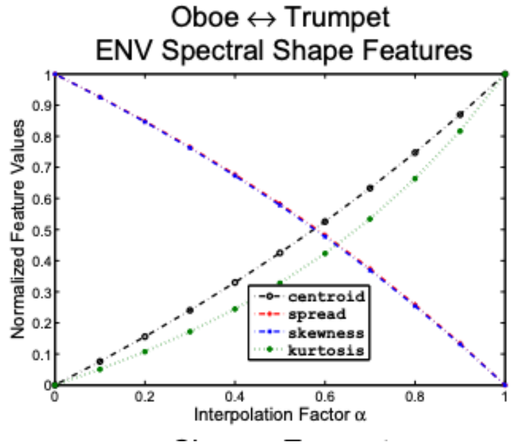


Illustration of two-dimensional timbre space with two sound objects depicted as the circle and the square and one intermediate sound object depicted as the square with rounded corners



$$S_{12} = \alpha S_1 + [1 - \alpha]\, S_2$$





(Caetano) Musical Instrument Sound Morphing Guided by Perceptually Motivated Features. *IEEE Transactions on Audio Speech and Language Processing 21(8), 2013*

- https://experiments.withgoogle.com/ai/sound-maker/view/

Timbre

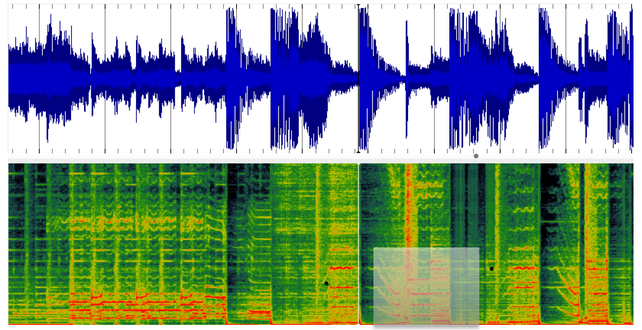
Song Identification (I)

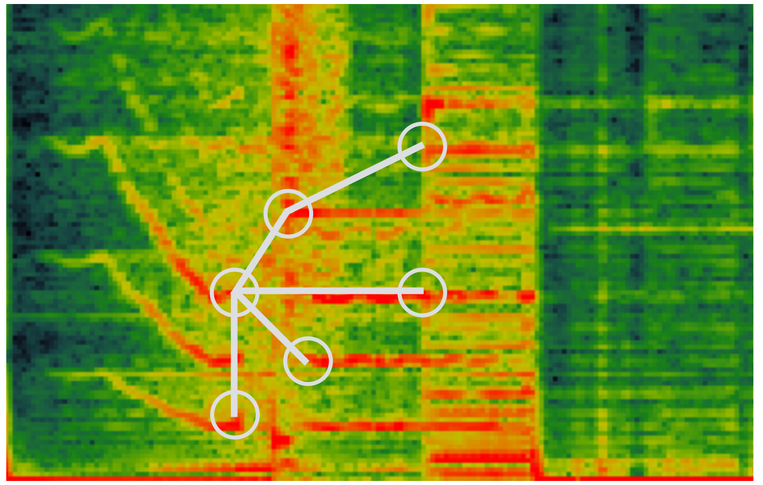
- Identify the song that is playing
- Identify it robustly and quickly
  - Search 30M (>>) songs as fast as possible
  - Find song based on 2s only
  - (for a business…) need industrial strength performance

Timbre

Song Identification (II)

- “Landmark” Approach / Fingerprinting
  - Find strongest spectral peaks (in time and frequency)
  - Make pattern of locally related peaks (like star constellation)
  - Have many patterns per second
  - Combination of patterns through time creates uniqueness -> musical fingerprint
  - Identify songs by matching patterns using hash tables (v. fast search)





Resources

General

- A large set of audio features for sound description (similarity and classification) in the CUIDADO project, *G. Peeters*, (2004)
- Instrument sound description in the context of MPEG-7, *G. Peeters, S. McAdams, P. Herrera*, ICMC (2000)
- Content processing of music audio signal, *F. Gouyon et al.*, Sound to sense, sense to sound – a state of the are in sound and music computing. Logos-Verlag, Berlin (2008).
- Automatic Classification of Pitched Musical Instrument Sounds. *Herrera-Boyer, P., Klapuri, A., & Davy, M.* In A. Klapuri & M. Davy (Eds.), Signal Processing Methods for Music Transcription (pp. 163–200). Springer. (2006)

Software

- The Timbre Toolbox, *Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S.* (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. The Journal of the Acoustical Society of America, 130(5), 2902–2916.
- MIR Toolbox, *Lartillot, O., & Toiviainen, P.* (2007). A Matlab Toolbox for Musical Feature Extraction from Audio. Proc of the 10th International Conference on Digital Audio Effects DAFx07, 1–8.
- Essentia, *Bogdanov, D., Wack, N., Emilia, G., Gulati, S., Herrera, P., Mayor, O., Roma, G., & Salamon, J.* (2013). Essentia: An Audio Analysis Library for Music Information Retrieval. ISMIR 2013

4.1. Timbre

.