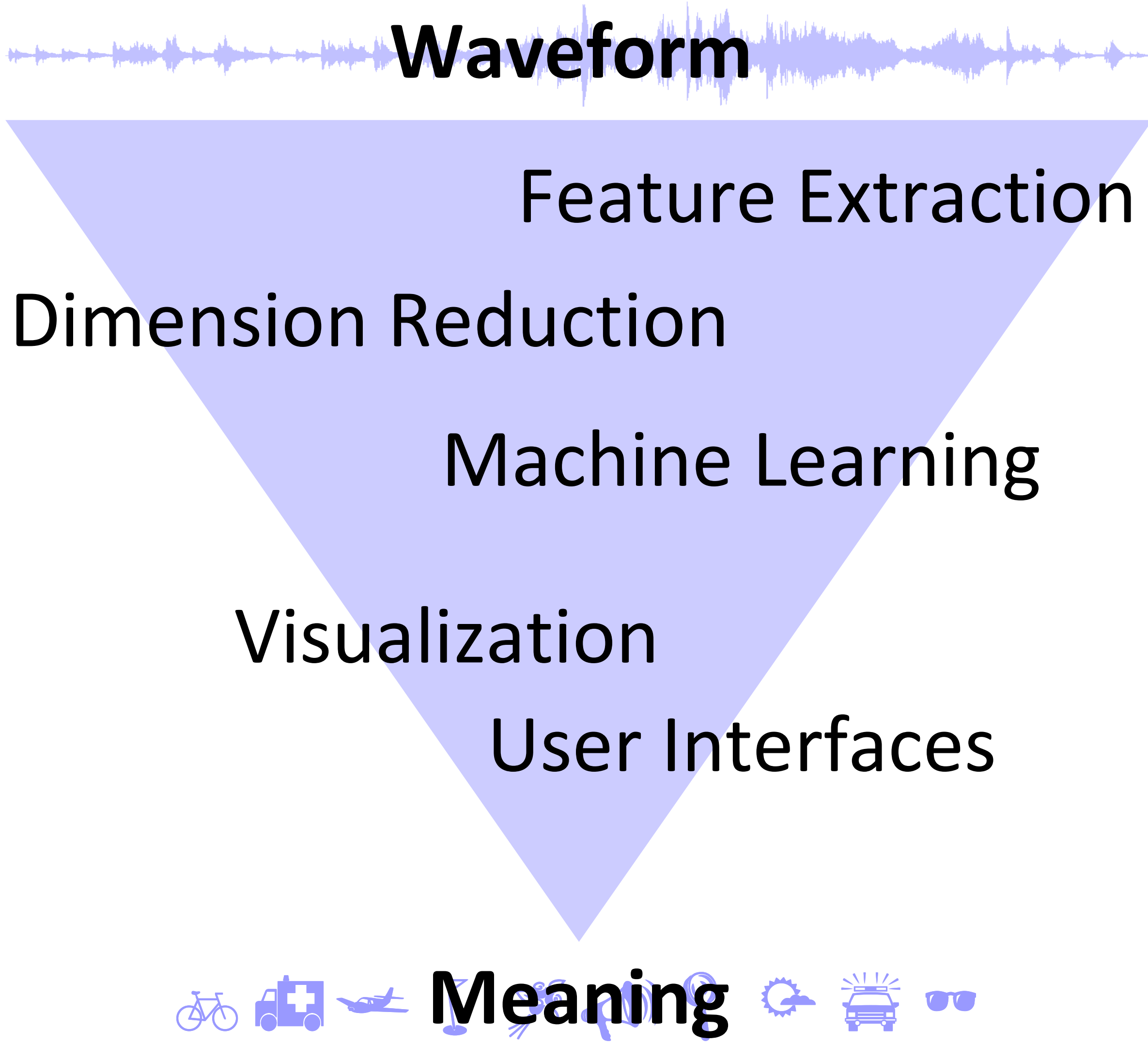
Do-It-Yourself Semantic Audio

Jörn Loviscach

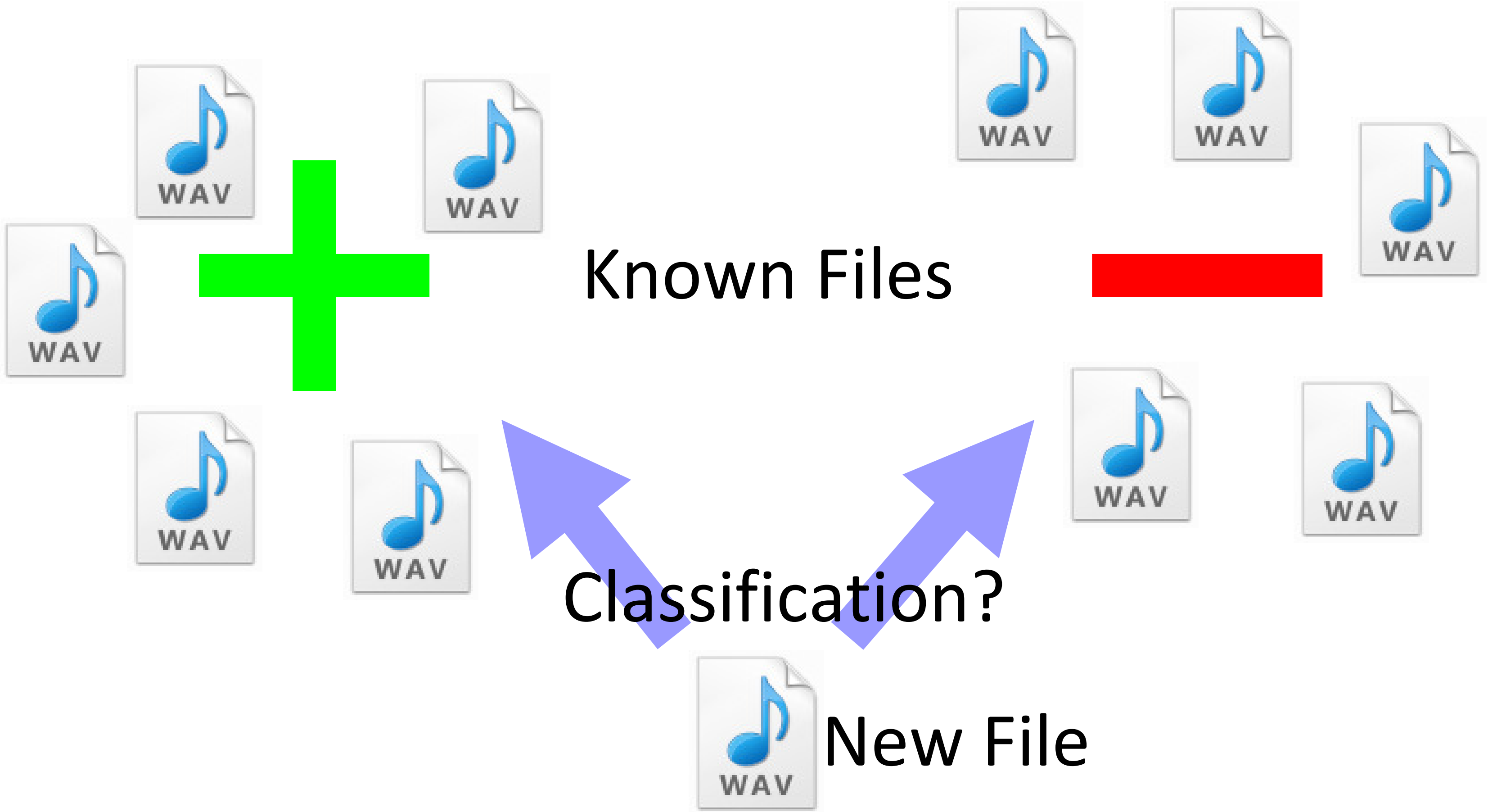
Fachhochschule Bielefeld, Germany
(Bielefeld University of Applied Sciences)

The Funnel Principle



Some Applications: Music Information Retrieval (1)

Find music similar to music that the listener likes

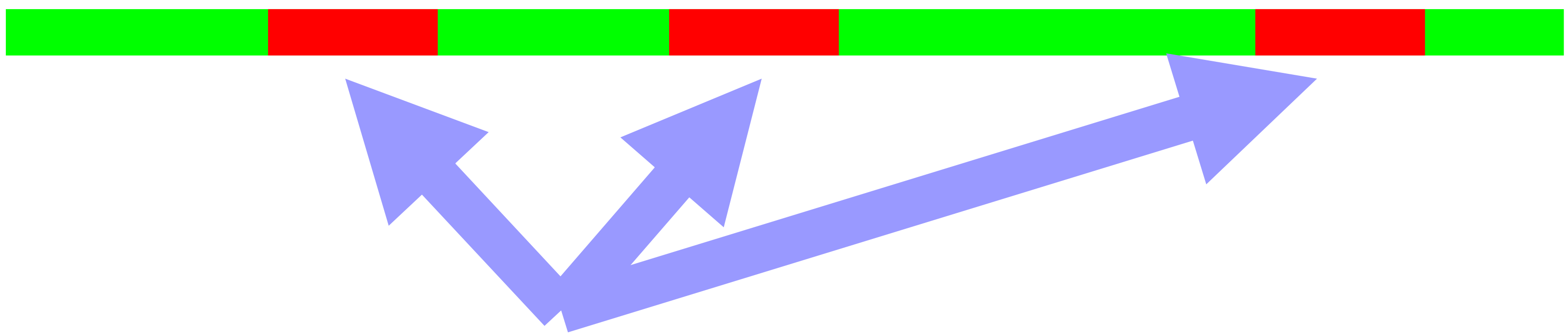


Some Applications: Music Information Retrieval (2)

Find music that fits to walking/jogging

Some Applications: Music Information Retrieval (3)

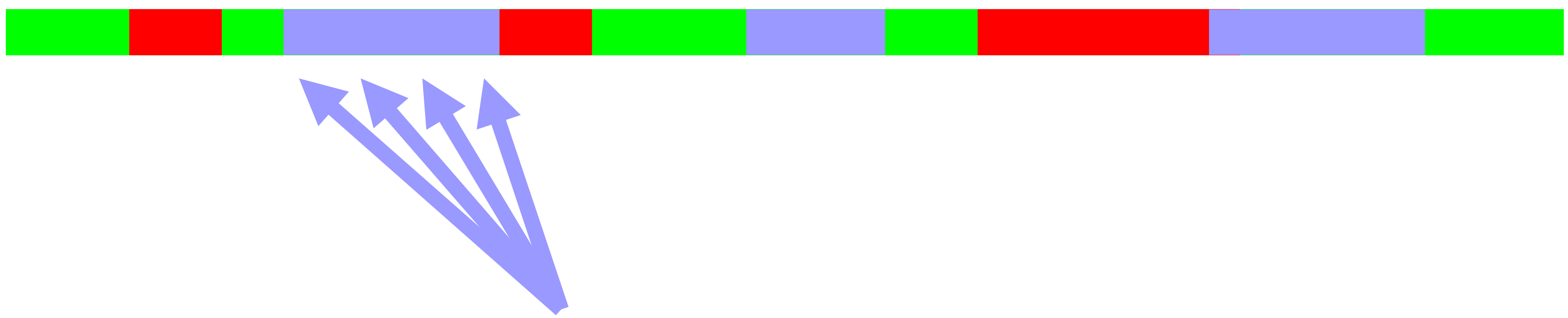
Extract the chorus of a song



Find the most prominent repeated part

Some Applications: Music Information Retrieval (4)

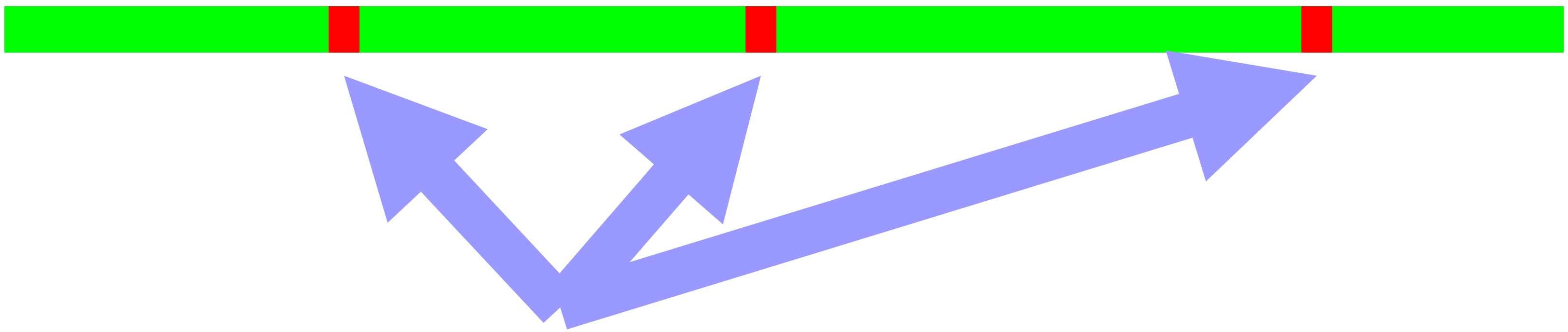
Segment radio archives: news, music, ads, etc.



Cluster temporal evolution and classify those clusters

Some Applications: Forensics

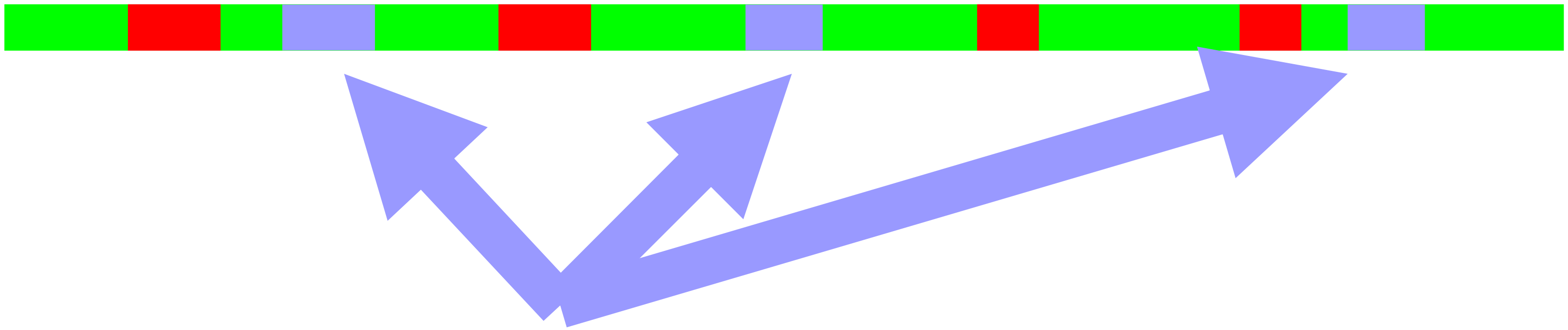
Detect gunshots in surveillance recordings



Find and classify
acoustic events

Some Applications: Language Learning

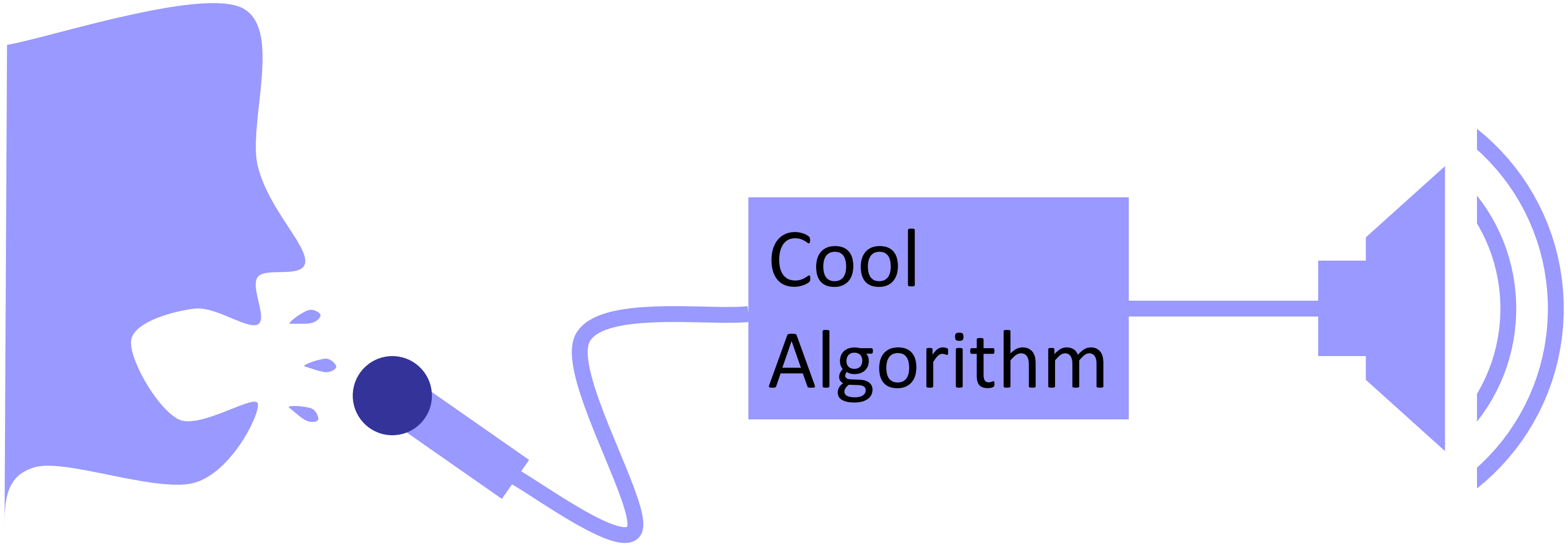
Identify the accent of a speaker



Recognize phonemes
and classify their timbre

Some Applications: Music Making

Control digital musical instruments acoustically



Objective of this Tutorial

Get going

- for free
- without C++ programming

Basic methods of

- Feature Extraction
- Machine Learning

Agenda

- The software landscape
- Basic feature extraction:
  - Sonic Visualiser
  - jAudio and Excel
- Feature extraction and machine learning:
  - jAudio and WEKA
  - MIRtoolbox in MATLAB®
- Real-time applications:
  - timbreID in Pure Data

Agenda

Time for Longer Questions

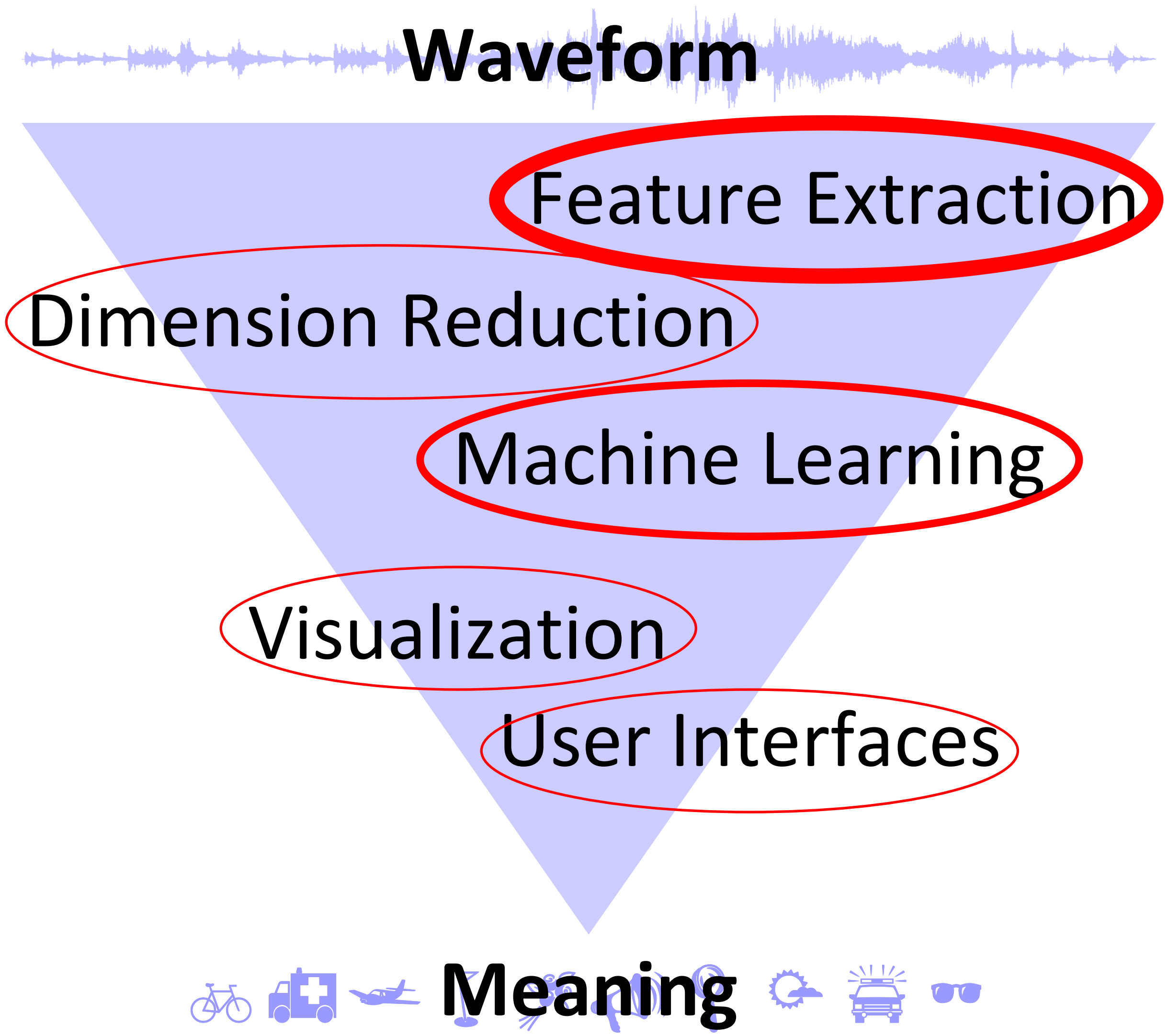
- The software landscape
- Basic feature extraction:
  - Sonic Visualiser
  - jAudio and Excel
- Feature extraction and machine learning:
  - jAudio and WEKA
  - MIRtoolbox in MATLAB®
- Real-time applications:
  - timbreID in Pure Data

Agenda

- The software landscape
- Basic feature extraction:
  - Sonic Visualiser
  - jAudio and Excel
- Feature extraction and machine learning:
  - jAudio and WEKA
  - MIRtoolbox in MATLAB®
- Real-time applications:
  - timbreID in Pure Data

Questions so far?

The Software Landscape: Scope



The Software Landscape: Offline vs. Real Time

- Offline processing

  Currently the typical mode

- Real-time processing

  Applications:

  – Score following & chord recognition for live music

  – Live control of digital musical instruments

The Software Landscape: Packaging

Many shapes and forms …

	Feature Extraction	Machine Learing	User Interface	Offline	Real-time	C/C++/Java	MATLAB®	Pure Data, Max/MSP	Stand-alone	Web Service	Active Development	Web Address
MARSYAS	✓	✓	⊘	✓	✓	✓	⊘	⊘	✓	⊘	✓	http://marsyas.info/
CLAM	✓	⊘	⊘	✓	✓	✓	⊘	⊘	⊘	⊘	✓	http://clam-project.org/
openSMILE	✓	✓	⊘	✓	✓	✓	⊘	⊘	✓	⊘	✓	http://opensmile.sourceforge.net/
LibXtract	✓	⊘	⊘	✓	⊘	✓	⊘	✓	✓	⊘	✓	http://sourceforge.net/projects/libxtract/
Aubio	✓	⊘	⊘	✓	⊘	✓	⊘	⊘	⊘	⊘	✓	http://aubio.org/
jAudio (part of jMIR)	✓	⊘	✓	✓	⊘	✓	⊘	⊘	✓	⊘	⊘	http://sourceforge.net/projects/jaudio/
Music-to-Knowledge (M2K)	✓	✓	⊘	✓	⊘	✓	✓	⊘	✓	⊘	⊘	http://www.music-ir.org/evaluation/m2k/
Maate	✓	⊘	⊘	✓	⊘	✓	⊘	⊘	⊘	⊘	⊘	http://lwn.net/2002/0321/a/maate.php3
MPEG-7 Audio Encoder	✓	⊘	⊘	✓	⊘	✓	⊘	⊘	✓	⊘	⊘	http://mpeg7audioenc.sourceforge.net/
MIRtoolbox	✓	✓	⊘	✓	⊘	⊘	✓	⊘	⊘	⊘	✓	https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox/mirtoolbox
MA Toolbox	✓	✓	⊘	✓	⊘	⊘	✓	⊘	⊘	⊘	⊘	http://www.ofai.at/~elias.pampalk/ma/
Computer Audition Toolbox (CATbox)	✓	✓	⊘	✓	⊘	⊘	✓	⊘	⊘	⊘	⊘	http://cosmal.ucsd.edu/cal/projects/CATbox/catbox.htm
MPEG-7 XM	✓	⊘	⊘	✓	⊘	⊘	✓	⊘	⊘	⊘	⊘	http://mpeg7.doc.gold.ac.uk/mirror/index.html
PsySound3	✓	✓	⊘	✓	⊘	⊘	✓	⊘	⊘	⊘	⊘	http://psysound.wikidot.com/
IPEM Toolbox	✓	✓	⊘	✓	⊘	⊘	✓	⊘	⊘	⊘	⊘	http://www.ipem.ugent.be/?q=node/27
Soundspotter	✓	✓	⊘	✓	✓	⊘	⊘	✓	⊘	⊘	✓	http://soundspotter.org/
timbreID	✓	✓	⊘	✓	✓	⊘	⊘	✓	⊘	⊘	✓	http://williambrent.conflations.com/pages/research.html
MuBu	✓	✓	✓	✓	✓	⊘	⊘	✓	⊘	⊘	✓	http://imtr.ircam.fr/imtr/MuBu
Chuck	✓	⊘	⊘	✓	✓	⊘	⊘	⊘	✓	⊘	✓	http://chuck.cs.princeton.edu/
Sonic Visualiser	✓	⊘	✓	✓	⊘	⊘	⊘	⊘	✓	⊘	✓	http://www.sonicvisualiser.org/
Sonic Annotator	✓	⊘	✓	✓	⊘	⊘	⊘	⊘	✓	⊘	✓	http://www.omras2.org/SonicAnnotator
Praat	✓	⊘	✓	✓	⊘	⊘	⊘	⊘	✓	⊘	✓	http://www.fon.hum.uva.nl/praat/
EchoNest	✓	✓	⊘	✓	⊘	⊘	⊘	⊘	⊘	✓	✓	http://developer.echonest.com/
MPEG-7 Audio Analyzer	✓	⊘	⊘	✓	⊘	⊘	⊘	⊘	⊘	✓	⊘	http://mpeg7lld.nue.tu-berlin.de/

Libraries for C/C++ or Java

Toolboxes for MATLAB®

Extensions for Pure Data and Max/MSP

Stand-alone solutions

Web Services

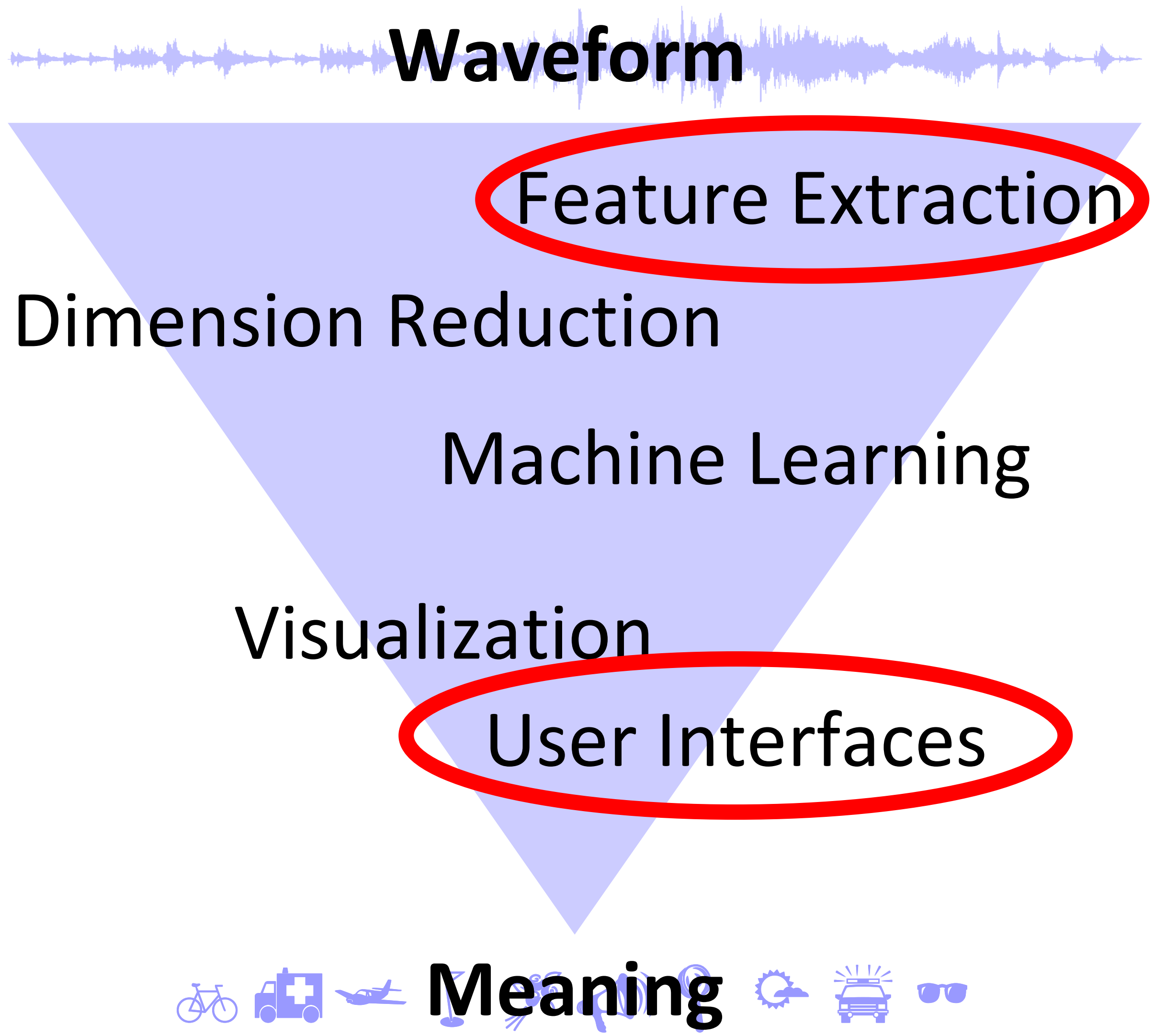
	Feature Extraction	Machine Learing	User Interface	Offline	Real-time	C/C++/Java	MATLAB®	Pure Data, Max/MSP	Stand-alone	Web Service	Active Development	Web Address
MARSYAS	✓	✓	⊘	✓	✓	✓	⊘	⊘	✓	⊘	✓	http://marsyas.info/
CLAM	✓	⊘	⊘	✓	✓	✓	⊘	⊘	⊘	⊘	✓	http://clam-project.org/
openSMILE	✓	✓	⊘	✓	✓	✓	⊘	⊘	✓	⊘	✓	http://opensmile.sourceforge.net/
LibXtract	✓	⊘	⊘	✓	⊘	✓	⊘	✓	✓	⊘	✓	http://sourceforge.net/projects/libxtract/
Aubio	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
jAudio (part of jMIR)	✓	⊘	✓	✓	⊘	✓	⊘	⊘	✓	⊘	⊘	http://sourceforge.net/projects/jaudio/
Music-to-Knowledge (M2K)	✓	✓	⊘	✓	⊘	✓	✓	⊘	✓	⊘	⊘	http://www.music-ir.org/evaluation/m2k/
Maate	✓	⊘	⊘	✓	⊘	✓	⊘	⊘	⊘	⊘	⊘	http://lwn.net/2002/0321/a/maate.php3
MPEG-7 Audio Encoder	✓	⊘	⊘	✓	⊘	✓	⊘	⊘	✓	⊘	⊘	http://mpeg7audioenc.sourceforge.net/
MIRtoolbox	✓	✓	⊘	✓	⊘	⊘	✓	⊘	⊘	⊘	✓	https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox/mirtoolbox
MA Toolbox	✓	✓	⊘	✓	⊘	⊘	✓	⊘	⊘	⊘	⊘	http://www.ofai.at/~elias.pampalk/ma/
Computer Audition Toolbox (CATbox)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]sd.edu/cal/projects/CATbox/catbox.htm
MPEG-7 XM	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]c.gold.ac.uk/mirror/index.html
PsySound3	✓	✓	⊘	✓	⊘	⊘	✓	⊘	⊘	⊘	⊘	http://psysound.wikidot.com/
IPEM Toolbox	✓	✓	⊘	✓	⊘	⊘	✓	⊘	⊘	⊘	⊘	http://www.ipem.ugent.be/?q=node/27
Soundspotter	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
timbreID	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
MuBu	✓	✓	✓	✓	✓	⊘	⊘	✓	⊘	⊘	✓	http://imtr.ircam.fr/imtr/MuBu
Chuck	✓	⊘	⊘	✓	✓	⊘	⊘	⊘	✓	⊘	✓	http://chuck.cs.princeton.edu/
Sonic Visualiser	✓	⊘	✓	✓	⊘	⊘	⊘	⊘	✓	⊘	✓	http://www.sonicvisualiser.org/
Sonic Annotator	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible].omras2.org/SonicAnnotator
Praat	✓	⊘	✓	✓	⊘	⊘	⊘	⊘	✓	⊘	✓	http://www.fon.hum.uva.nl/praat/
EchoNest	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	⊘	✓	✓	http://developer.echonest.com/
MPEG-7 Audio Analyzer	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	⊘	✓	⊘	http://mpeg7lld.nue.tu-berlin.de/

Agenda

- The software landscape
- Basic feature extraction:
  - Sonic Visualiser
  - jAudio and Excel
- Feature extraction and machine learning:
  - jAudio and WEKA
  - MIRtoolbox in MATLAB®
- Real-time applications:
  - timbreID in Pure Data

Questions so far?

Sonic Visualiser



Sonic Visualiser

- Manual and automated markup
- Many feature extractors available;
  install in C:\Program Files (x86)\Vamp Plugins
- Great for experiments with feature extraction
- Things to see and try:
  - Details about current position of mouse pointer
  - Draw musical notes
  - Align timelines of two versions
    of a recording (plug-in)

Male/Female Segmentation

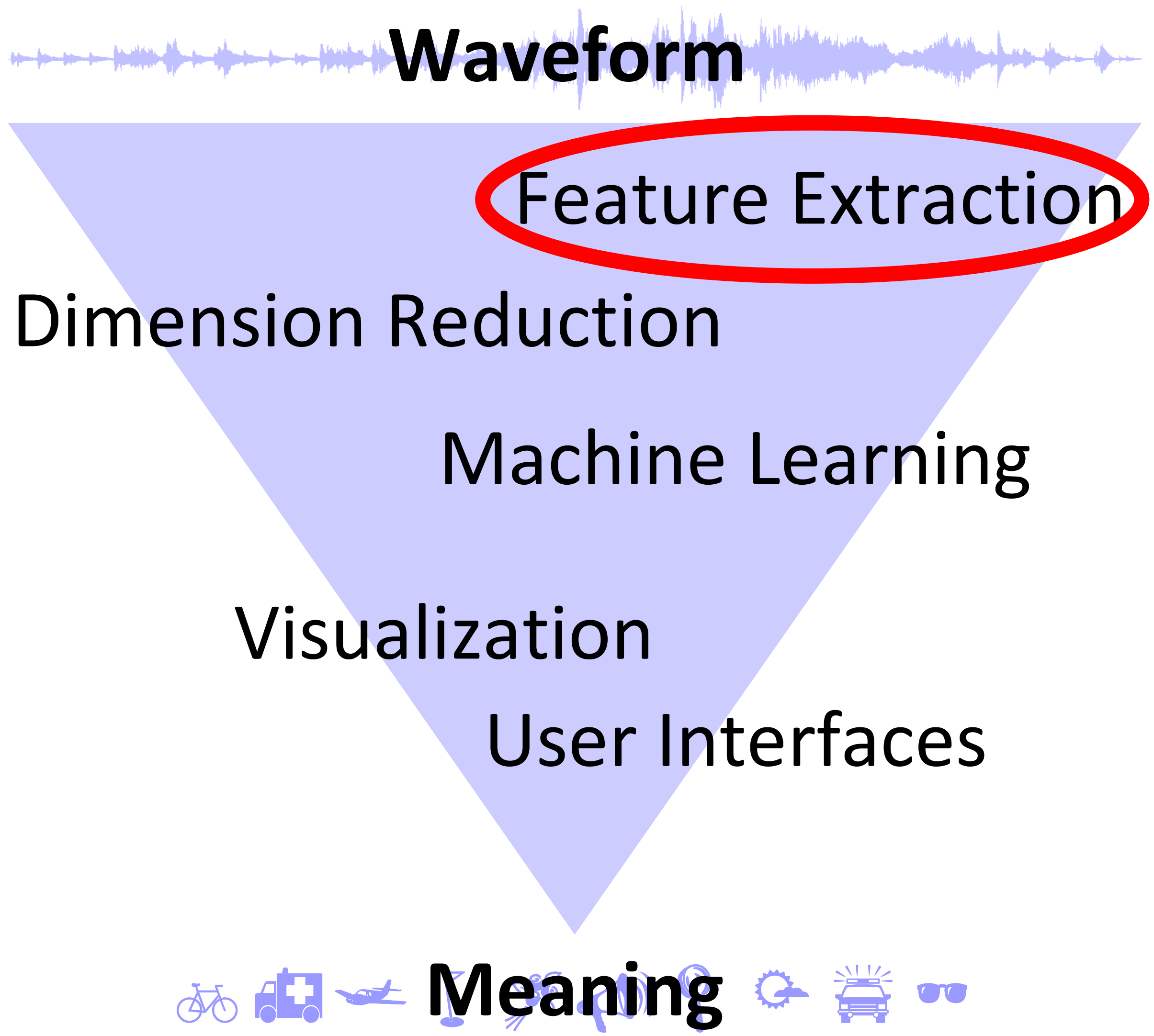
- Add new pane; add spectrogram
- Window: 32,768 samples; vert. axis logarithmic
- Add new time instants layer
- Add markers
- Plot type: segmentation
- Name markers (cross tool or edit layer data)
- Edit markers if needed
- Export annotation layer

Agenda

- The software landscape
- Basic feature extraction:
  - Sonic Visualiser
  - jAudio and Excel
- Feature extraction and machine learning:
  - jAudio and WEKA
  - MIRtoolbox in MATLAB®
- Real-time applications:
  - timbreID in Pure Data

Questions so far?

jAudio and Excel



jAudio: the Program

- Feature extractor
- Graphical user interface and command line
- Java-based
- Multi-threaded
- Batch processing (add multiple files at once!)
- Export e.g. as ACE (XML-based);
  nice for Excel

jAudio: Catches

- Install as admin
- Override standard heap size:
  No double-click to start, rather
  `java -Xmx1024M -jar jAudio.jar`
  in the directory of the jar. (Batch file!)
- No ä or é in audio file names:
  XML output broken
- XML and ARFF: cleartext. Huge files!
  Export as few values as possible.

Sorting Files by Loudness

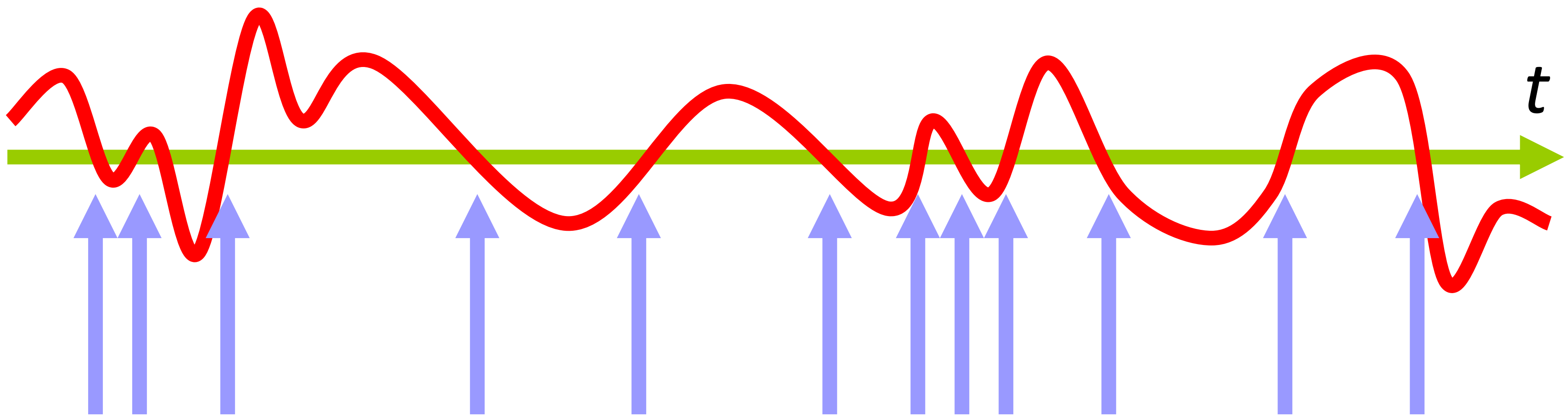
- Set paths for output files
- Do not export standard deviation (Alter Aggregators, click Save!)
- For each file, extract overall mean of root mean square
- Import into Microsoft Excel
- Sort and plot

Sorting Sounds by Brightness

- These are different ways of measuring brightness:
  - Number or rate of zero crossings
  - Spectral centroid
  - Spectral rolloff point
- jAudio: for each file, extract overall mean
- Import into Microsoft Excel
- Sort and/or plot (x = item number)

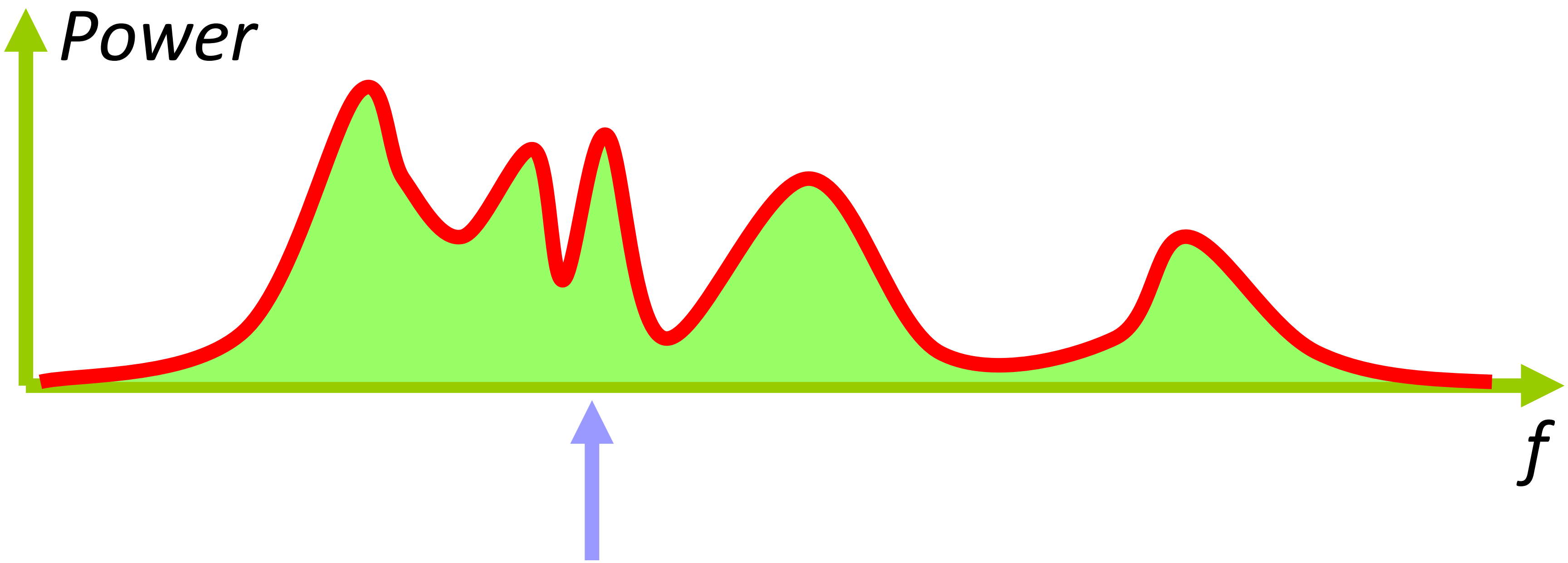
Zero Crossings

- Number or rate of sign changes
- Related to frequency and noise content
- Independent of volume
- Issue: sensitive to noise and harmonics



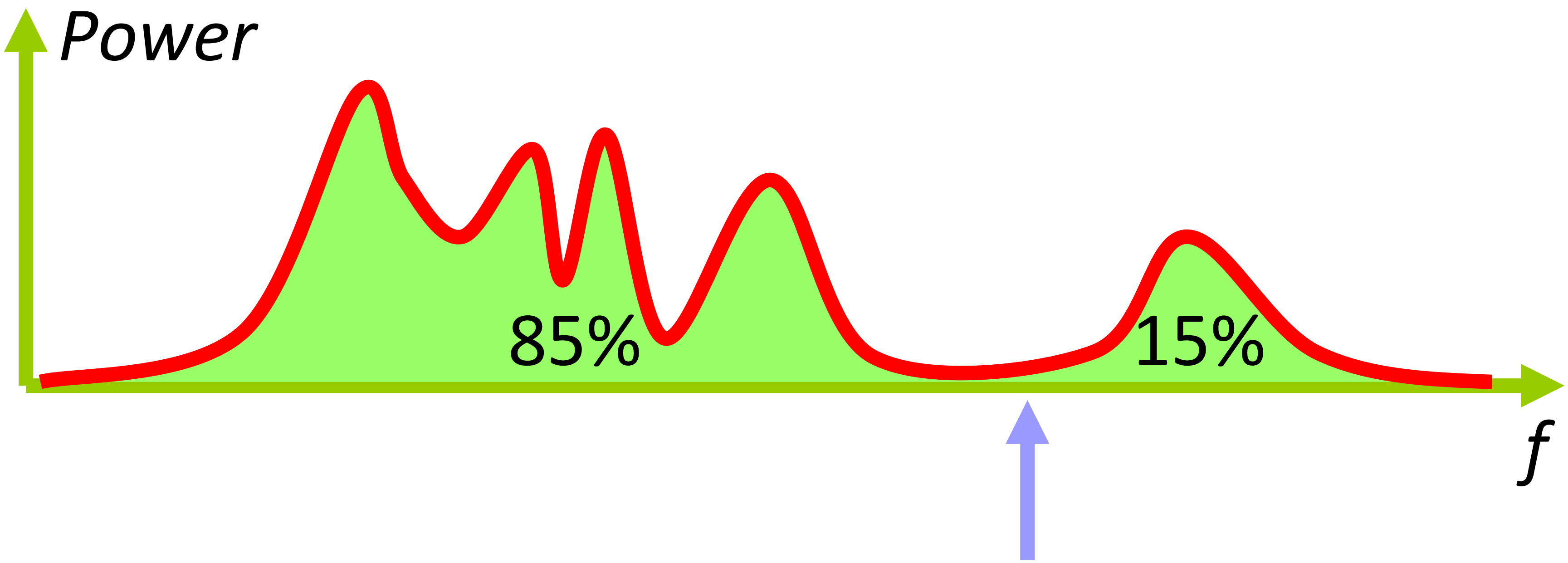
Spectral Centroid

- Mean frequency (center of mass) of the power spectrum (linear or log freq.)
- Independent of volume (if √Power)



Spectral Rolloff Point

- Determine the frequency that divides the audio power 85:15 (for instance)
- Independent of volume (if √Power)
- Fluctuating with empty spectral regions

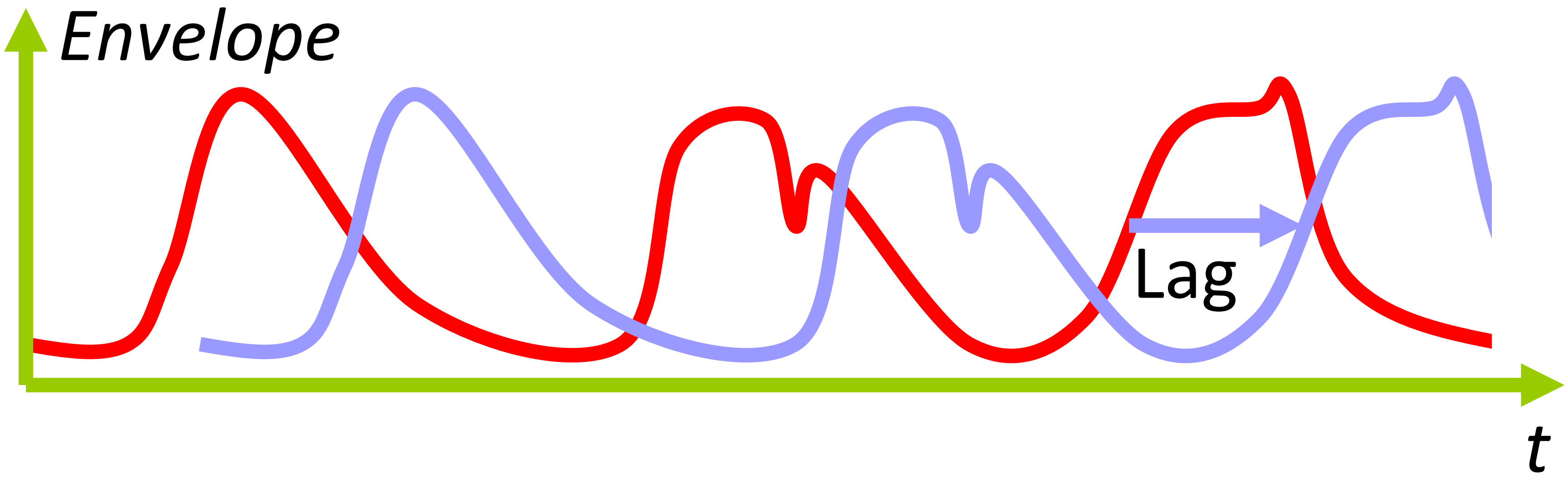


Sorting Music by Tempo

- Demos with Sonic Visualiser:
  - Note onsets
  - Beat and bar tracker
- jAudio:
  for each file, extract mean
  of strongest beat
- Import into Microsoft Excel
- Sort and/or plot (x = item number)

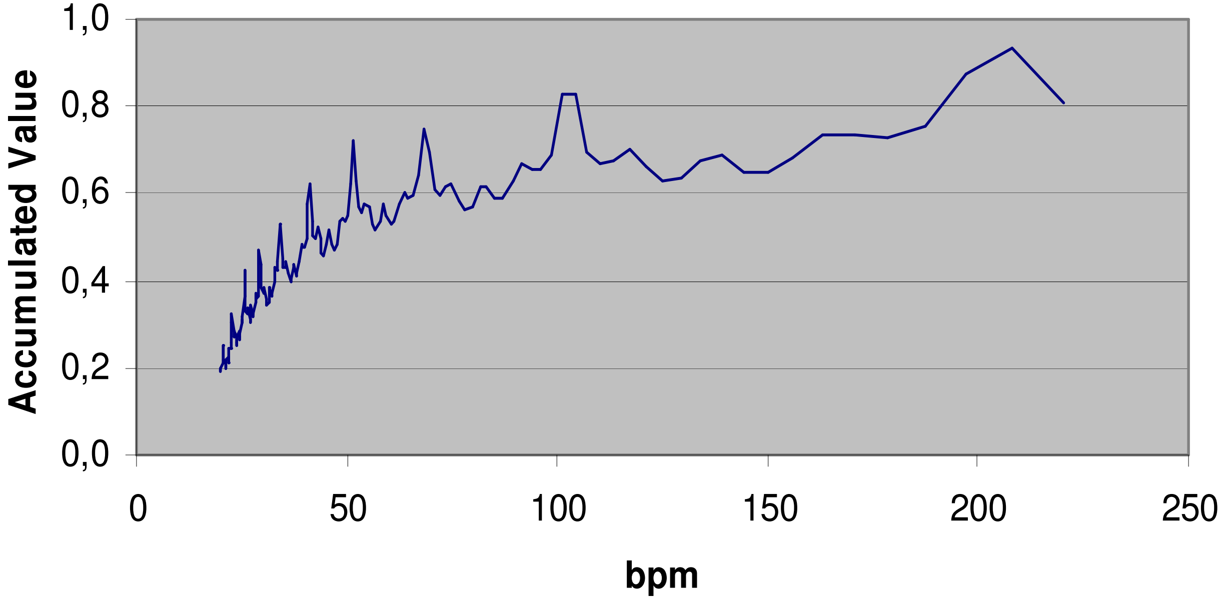
Strongest Beat

- Compute envelope
- Compute autocorrelation
- Return inverse of time lag
  of maximum autocorrelation (except 0)



Strongest Beat

- Issue with ambiguity:
  jAudio picks the maximum histogram bin
- Could improve that in Excel
  by extracting the full histogram

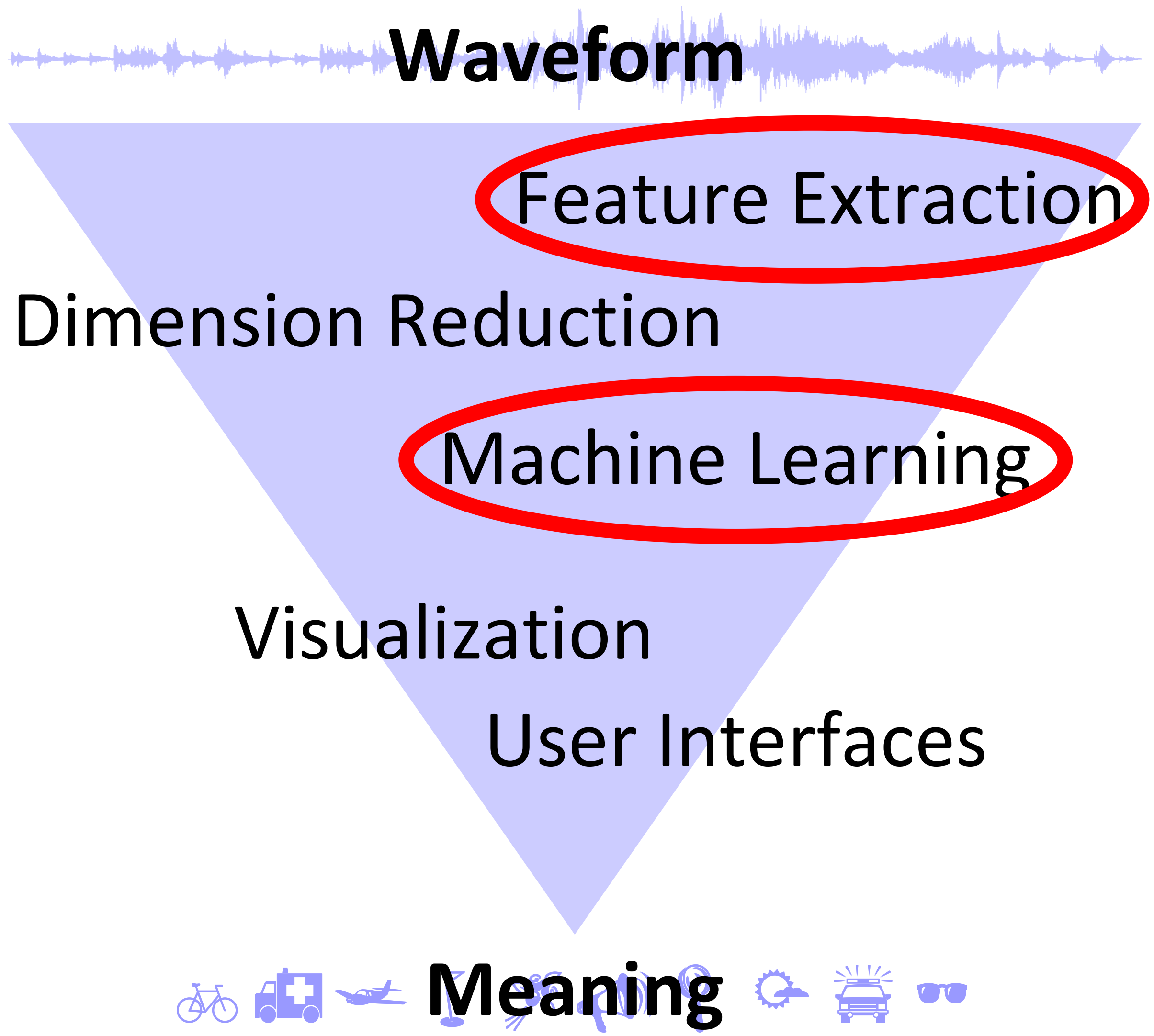


Agenda

- The software landscape
- Basic feature extraction:
  - Sonic Visualiser
  - jAudio and Excel
- Feature extraction and machine learning:
  - jAudio and WEKA
  - MIRtoolbox in MATLAB®
- Real-time applications:
  - timbreID in Pure Data

Questions so far?

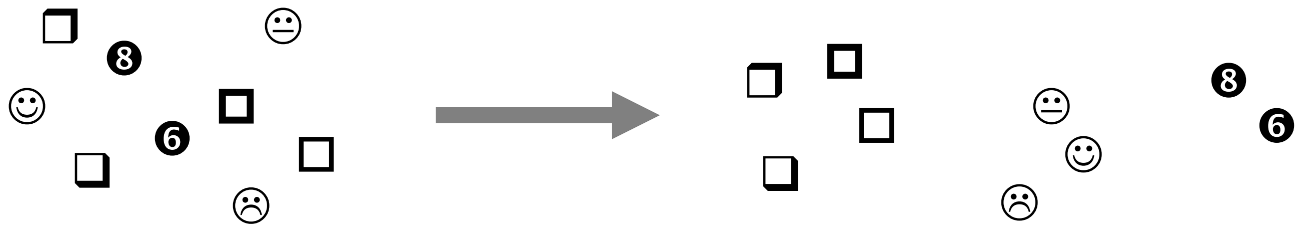
jAudio and WEKA



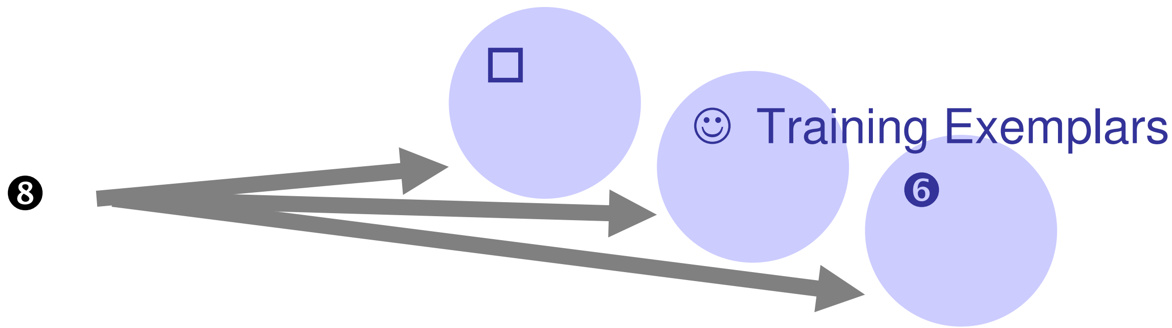
WEKA: the Program

Huge collection of machine learning algorithms

- Clustering: unsupervised machine learning



- Classification: supervised machine learning



WEKA: the Program

- Great for experiments
- ARFF: Plaintext file format for input data, one of the two formats written by jAudio
- Java-based
- In RunWeka.ini:
  `maxheap=512m`

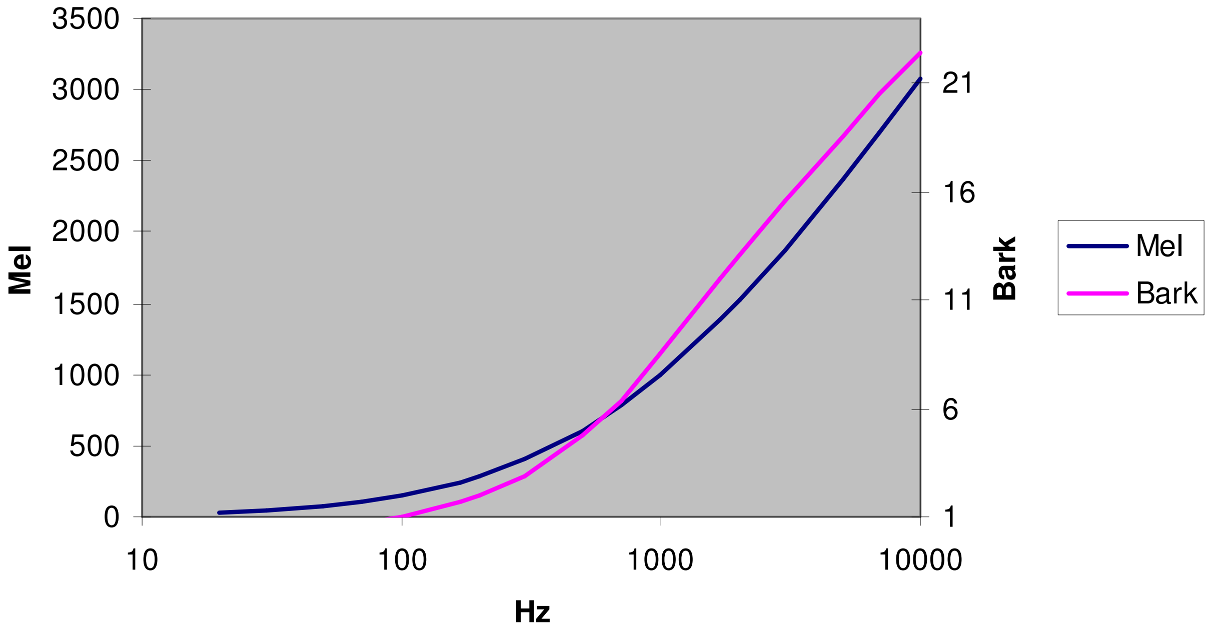
Clustering Sounds by Similarity

- Demo: vowel sounds
- jAudio: extract MFCC averages
- Export as ARFF (change file extension!)
- Import into WEKA Explorer: Preprocess
- Retain only the means of MFCCs 1…12
- Cluster:
  - Store clusters for visualization
  - Visualize cluster assignments

MFCCs: Mel-Frequency Cepstral Coefficients (1)

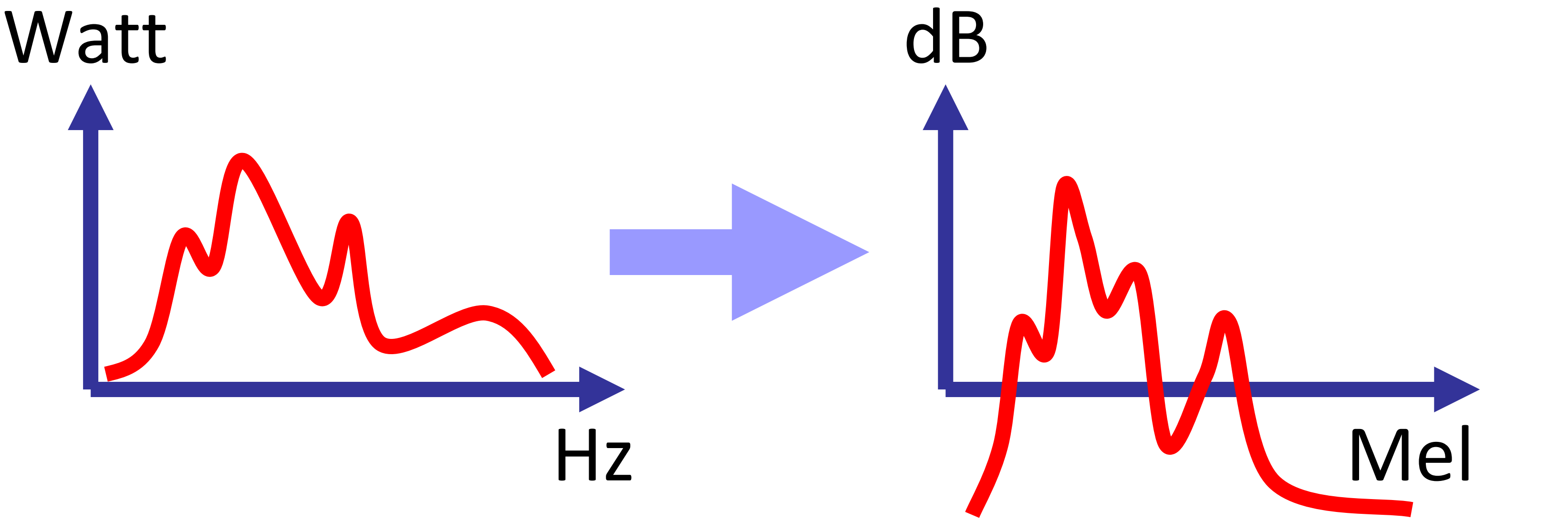
Rough idea of what the ear sends to the brain for one single moment of time

- Step 1: Short-time spectrum in perceived frequencies



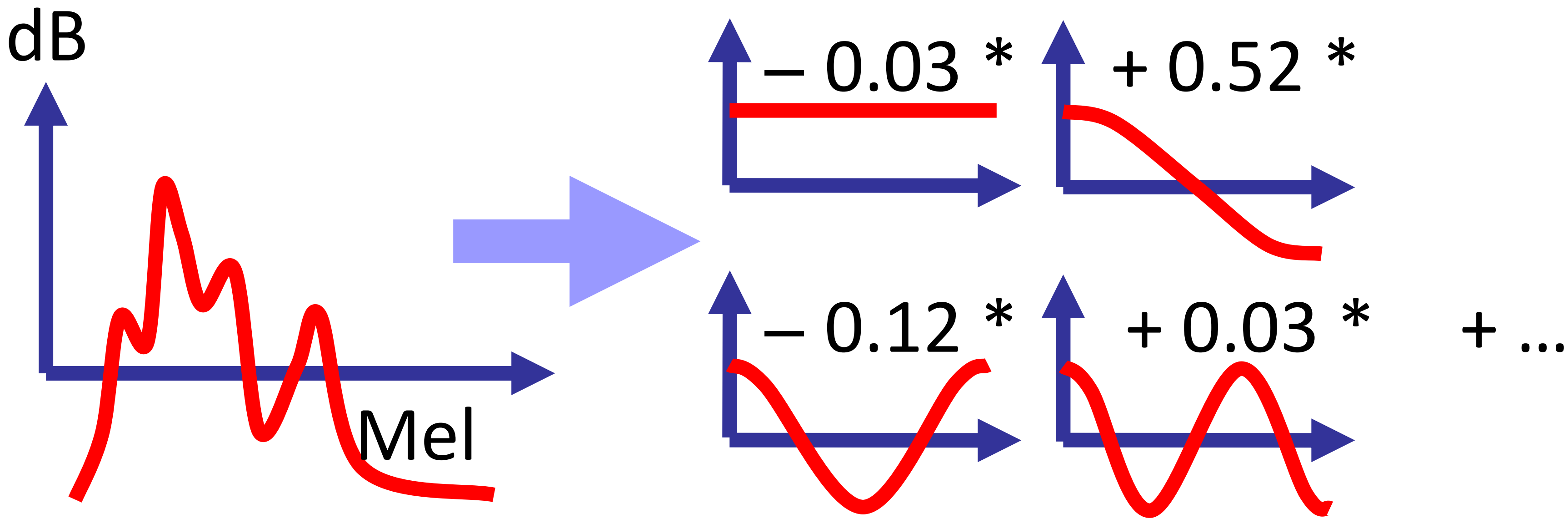
MFCCs: Mel-Frequency Cepstral Coefficients (2)

- Step 2: Compute approximate perceived loudness: log of power
- Intermediate result: spectrum as perceived



MFCCs: Mel-Frequency Cepstral Coefficients (3)

- Step 3: Describe the overall shape of this spectrum
- Do this through a mixture of cosine shapes
- MFCCs = the amounts of the different cosines

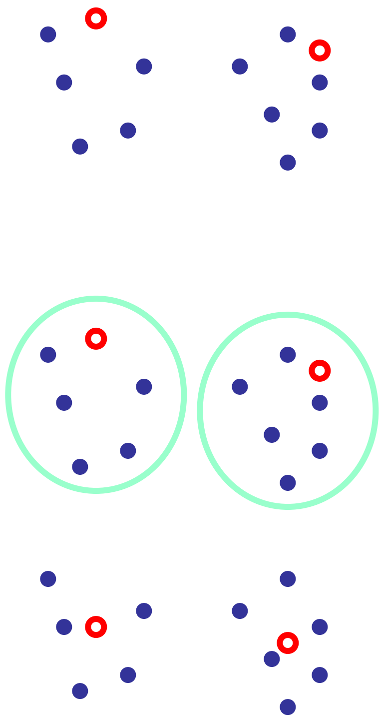


MFCCs: Mel-Frequency Cepstral Coefficients (4)

- Demo with Sonic Visualizer
- MFCC 0 ist just the audio level:
  Discard it to be independent of level
- Fine structure of spectrum is ignored
- What MFCCs are not designed to do:
  - Tell different fundamental frequencies apart
  - Distinguish harmonic/inharmonic/noise

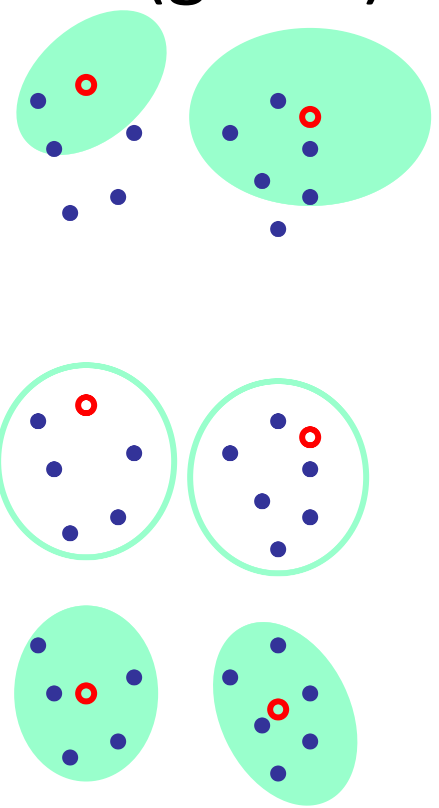
Clustering: k-Means

- Input: data points, number of clusters (guess)
- Pick random centers for clusters
- Iterate:
  - Assign each data point to the nearest center
  - New center = centroid of all points assigned
- Output: classification and centers



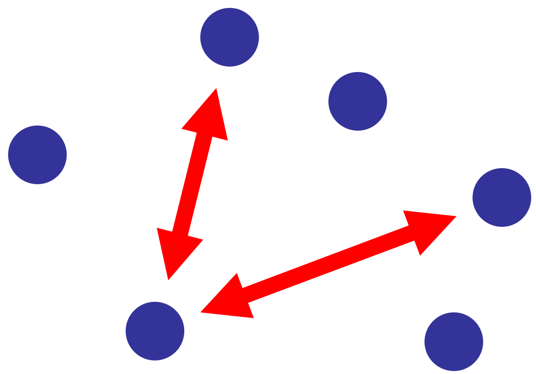
Clustering: Expectation Maximization (EM)

- Input: data points, number of clusters (guess)
- Pick random centers/sizes for clusters
- Iterate:
  - Assign each data point to the most probable center
  - New center/size according to points assigned
- Output: classification, centers, sizes



Clustering: Caveats

- Metric structure ≈ perception?



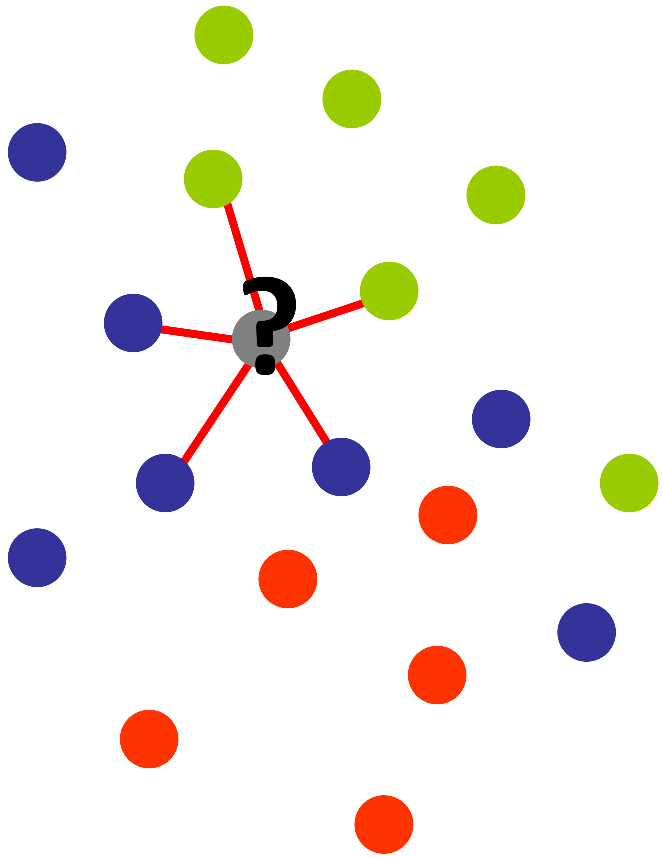
- Are all data dimensions of the right scale?
  - Weka: Visualize All
  - Weka: Standardize, Math Expression, …
- Vital when combining different features

Music Classification

- jAudio: extract MFCC averages
- Add to ARFF file:
  - @ATTRIBUTE class {classical, jazz, pop, rock}
  - Class of each file
- Import into WEKA Explorer
- Classify
- Visualize classifier errors

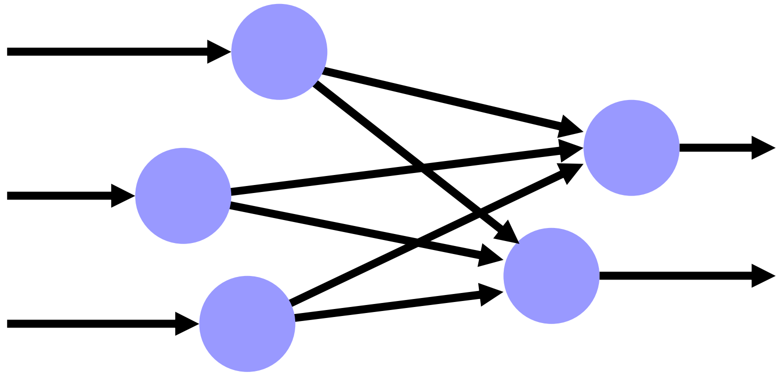
Classification: k Nearest Neighbors (kNN)

- Input:
  - Classified exemplars
  - The number k
  - The item x to be classified
- Find the k exemplars nearest to x
- Vote by majority among them

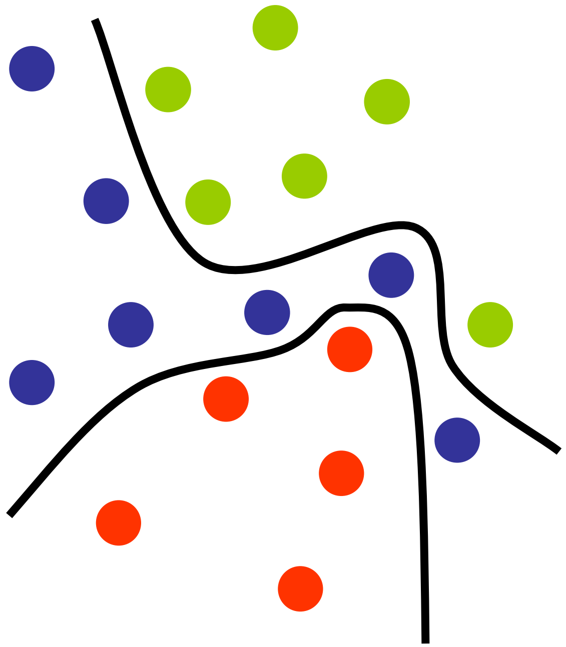


Classification: Zoo of Methods

- Neural Networks



- Support-Vector Machines



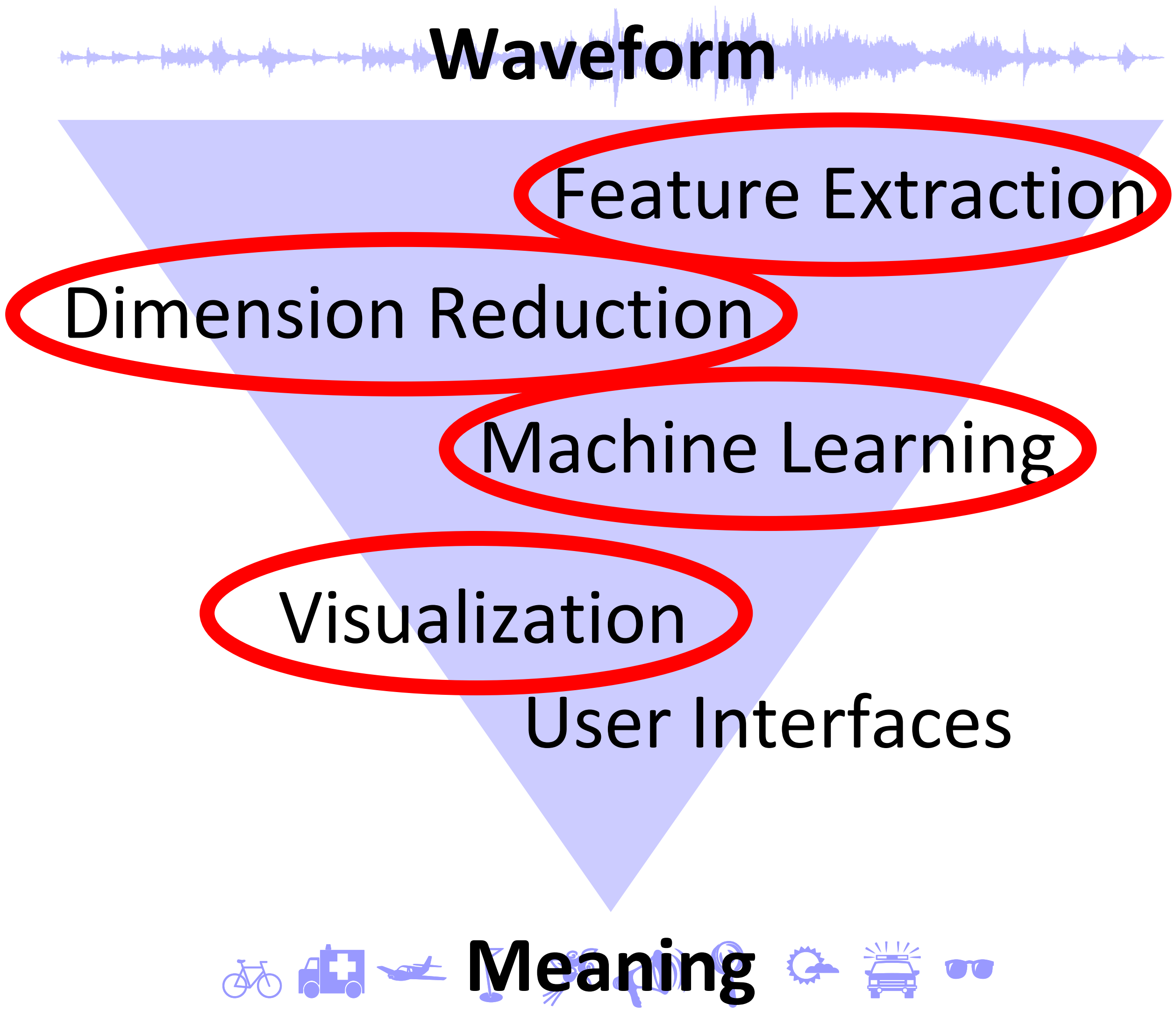
- and dozens more

Agenda

- The software landscape
- Basic feature extraction:
  - Sonic Visualizer
  - jAudio and Excel
- Feature extraction and machine learning:
  - jAudio and WEKA
  - MIRtoolbox in MATLAB®
- Real-time applications:
  - timbreID in Pure Data

Questions so far?

MIRtoolbox in MATLAB®



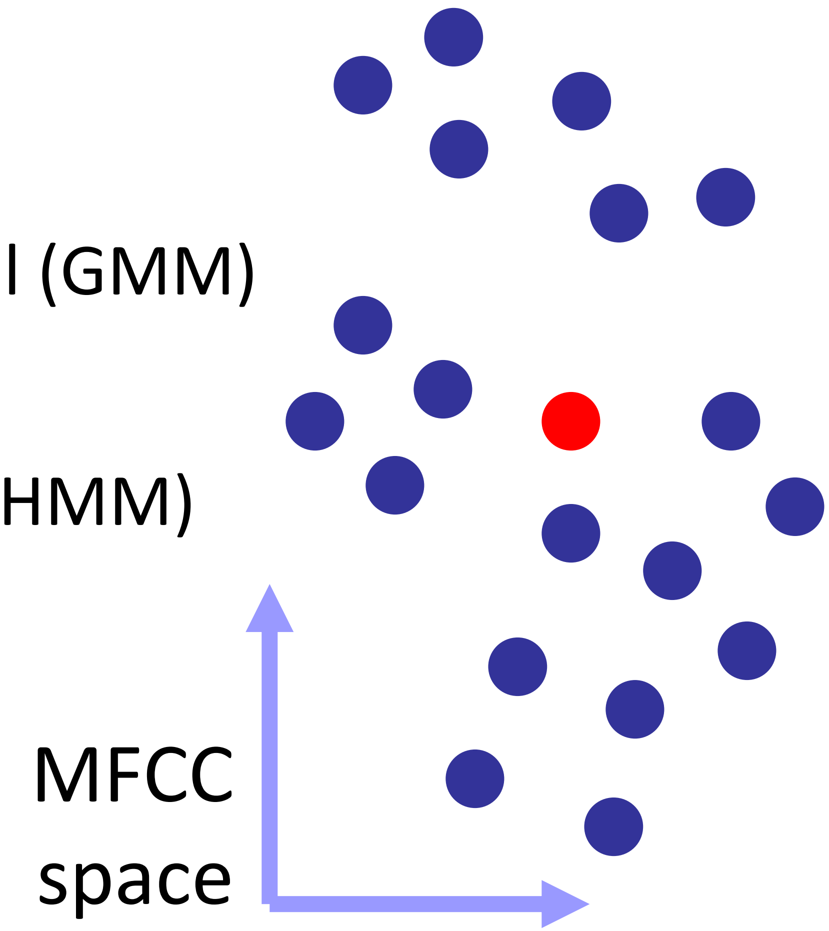
MIRtoolbox: the software

- All in one well-designed package,
  great for experimentation:
  - Low-level features
  - Dimension reduction
  - Machine Learning
- Requires MATLAB®, which is costly
- Slower than Java or C++,
  even though intermediate results are reused

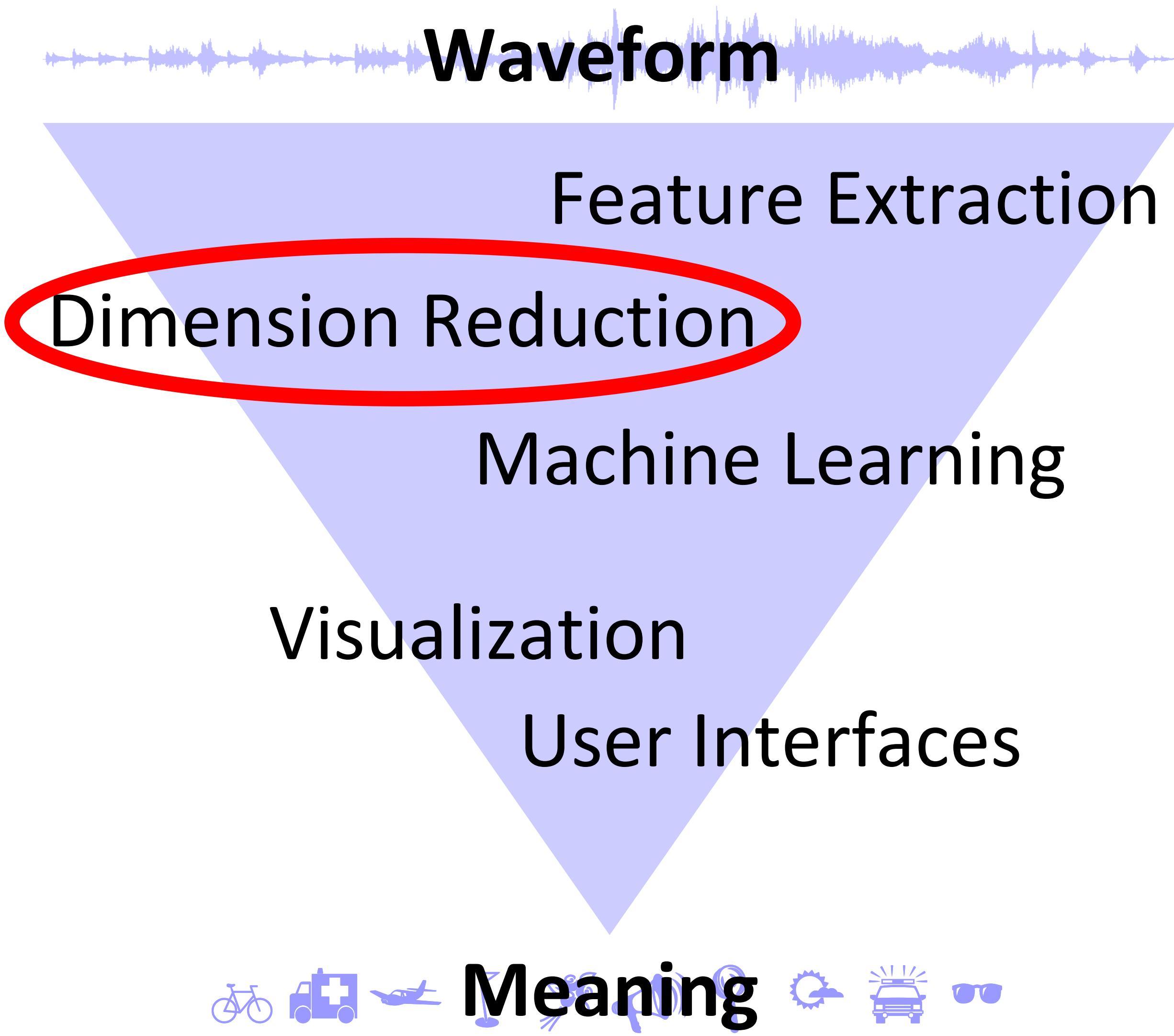
Music Classification: Improvements (1)

No average of MFCCs; better statistical model

- Ignoring time order:
  - k-Means
  - Gaussian Mixture Model (GMM)
- With time oder:
  - Hidden Markov Model (HMM)



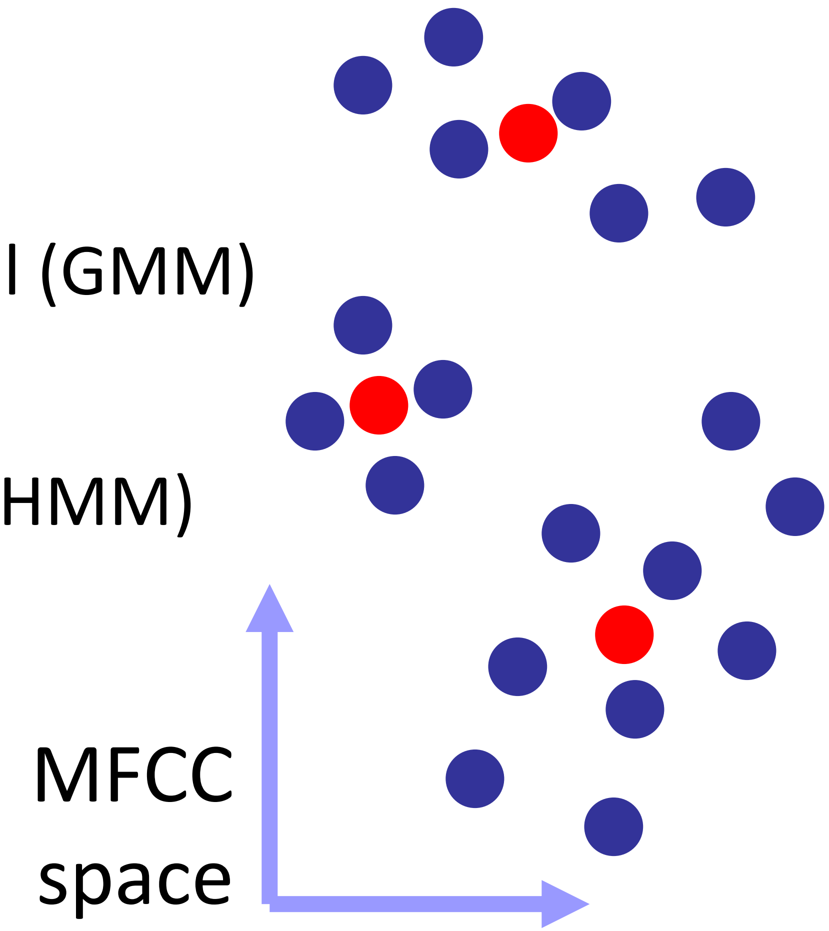
Music Classification: Improvements (2)



Music Classification: Improvements (3)

Not mean of MFCCs, but statistical model

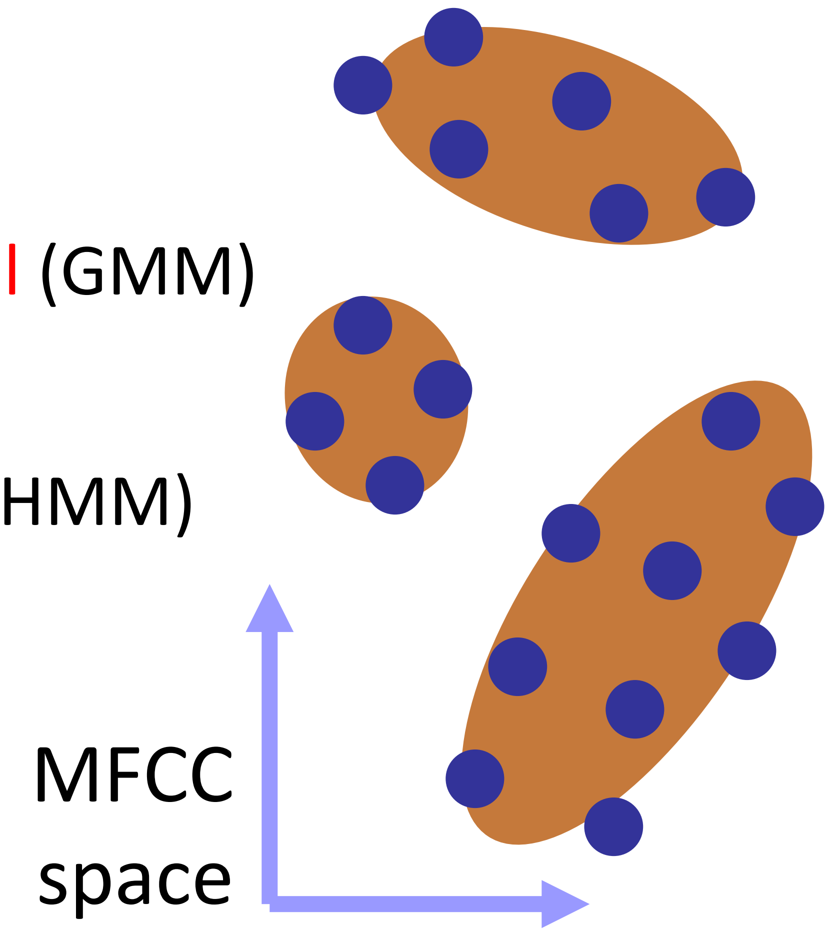
- Ignoring time order:
  - k-Means
  - Gaussian Mixture Model (GMM)
- With time oder:
  - Hidden Markov Model (HMM)



Music Classification: Improvements (4)

Not mean of MFCCs, but statistical model

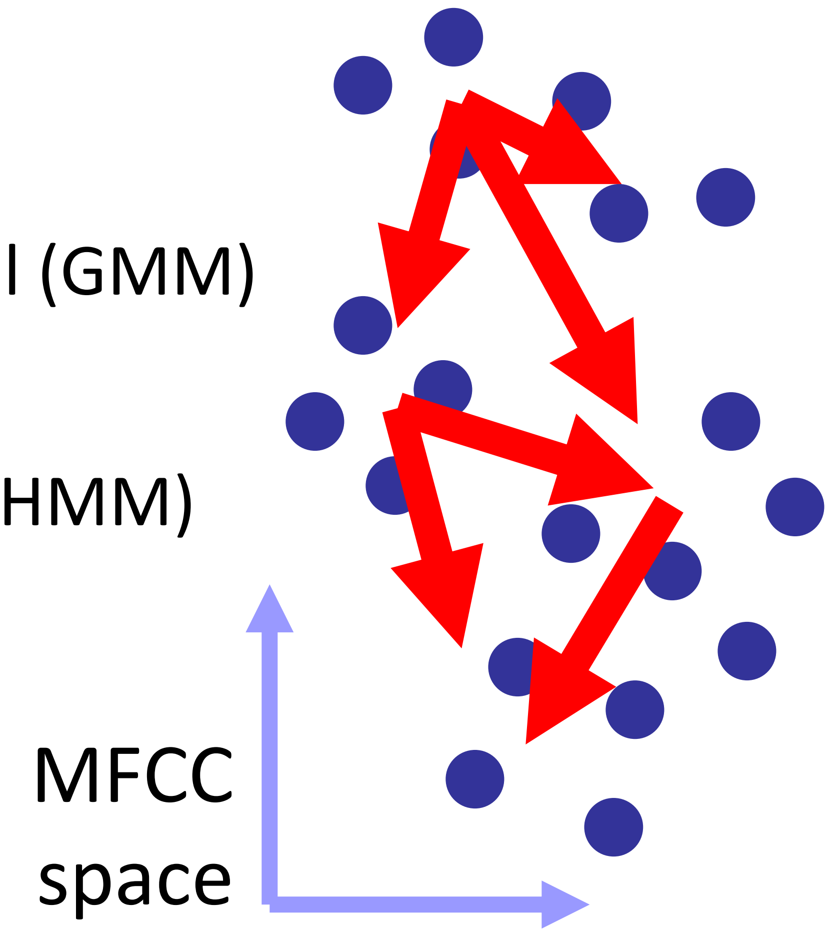
- Ignoring time order:
  - k-Means
  - Gaussian Mixture Model (GMM)
- With time oder:
  - Hidden Markov Model (HMM)



Music Classification: Improvements (5)

Not mean of MFCCs, but statistical model

- Ignoring time order:
  - k-Means
  - Gaussian Mixture Model (GMM)
- With time oder:
  - Hidden Markov Model (HMM)



MIRtoolbox in Action

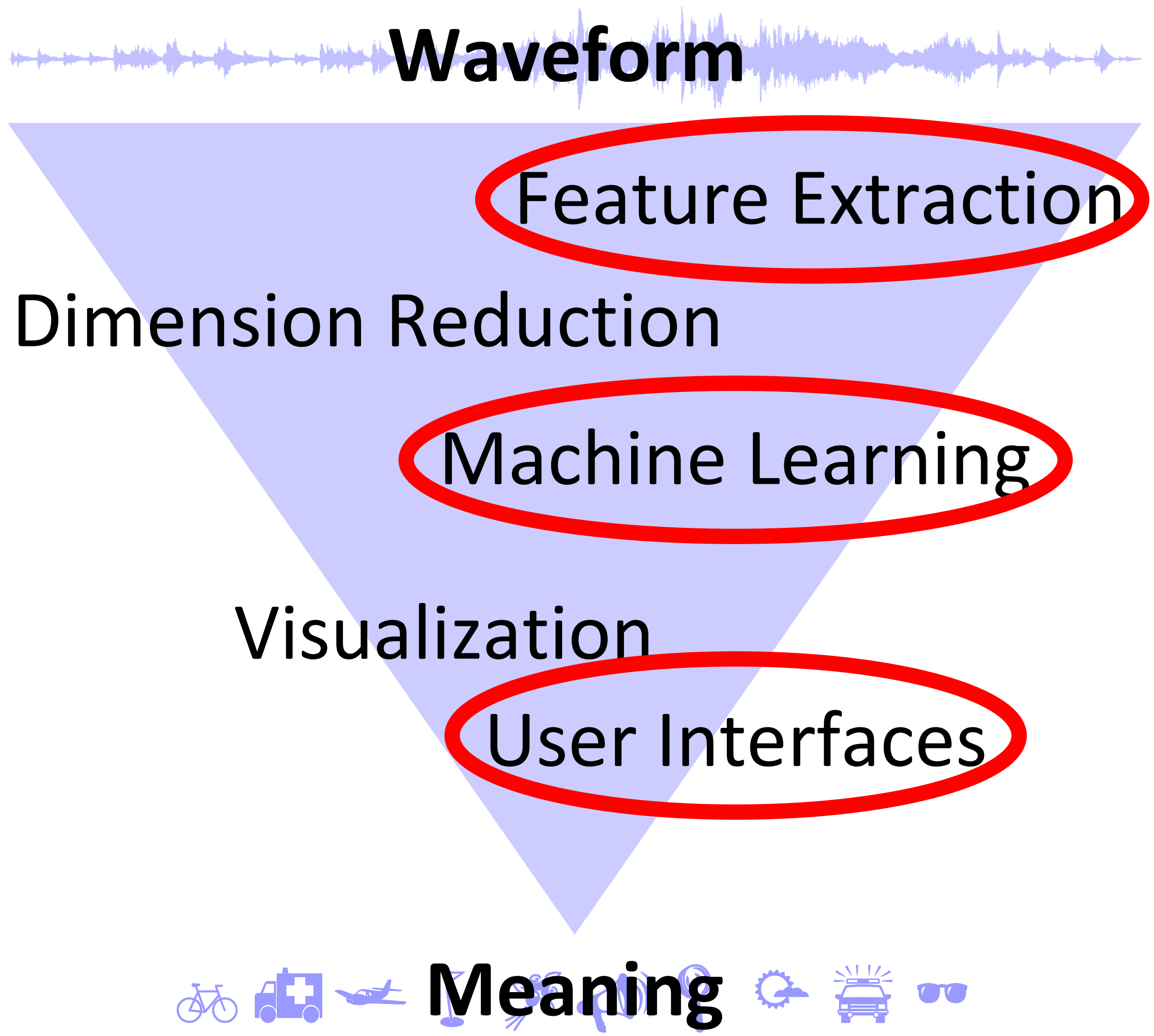
- Classify audio files by music genre
- Training set, test set:
  add prefixes to the files, e.g., p, r, j, c
- Extract features,
  condense by GMM,
  classify by Bayes

Agenda

- The software landscape
- Basic feature extraction:
  - Sonic Visualiser
  - jAudio and Excel
- Feature extraction and machine learning:
  - jAudio and WEKA
  - MIRtoolbox in MATLAB®
- Real-time applications:
  - timbreID in Pure Data

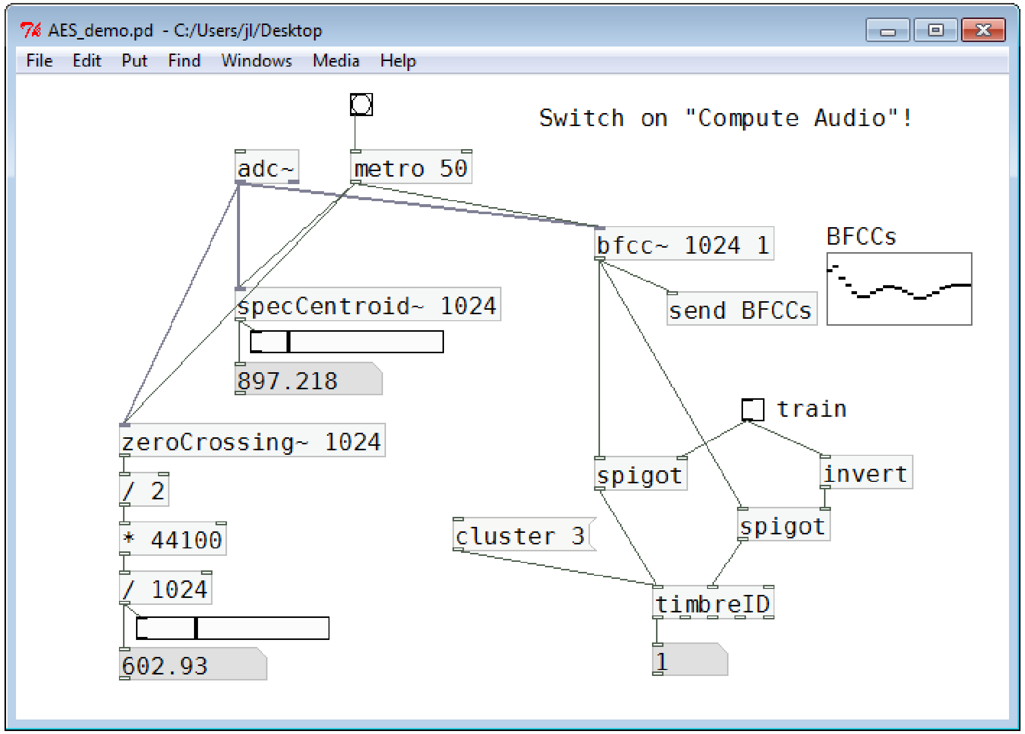
Questions so far?

timbreID in PureData



timbreID in Action

- Low-level features
- k-NN classification
- Clustering of training exemplars

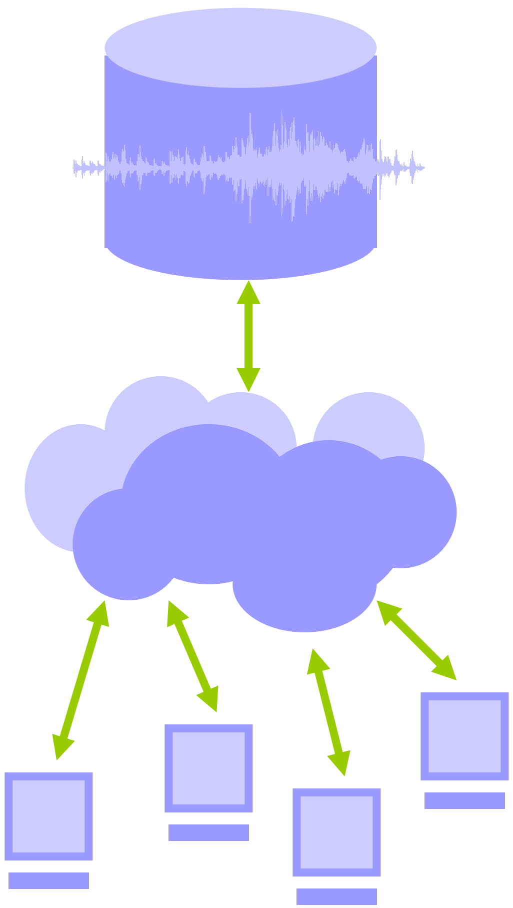


- Generate/control MIDI data, audio signals, …

Outlook

Outlook: Semantic Audio via the Internet

- EchoNest: Web Service for Music Information Retrieval
- Collect data from the users
- Keep waveforms (large, expensive, sensitive) away from the end user
- Mashups of Web Services?
- Real time, too??



Thank you!

www.j3L7h.de

Questions?