



Audio Mining with emphasis on Music Genre Classification

By Anders Meng, IMM
www.imm.dtu.dk/~am





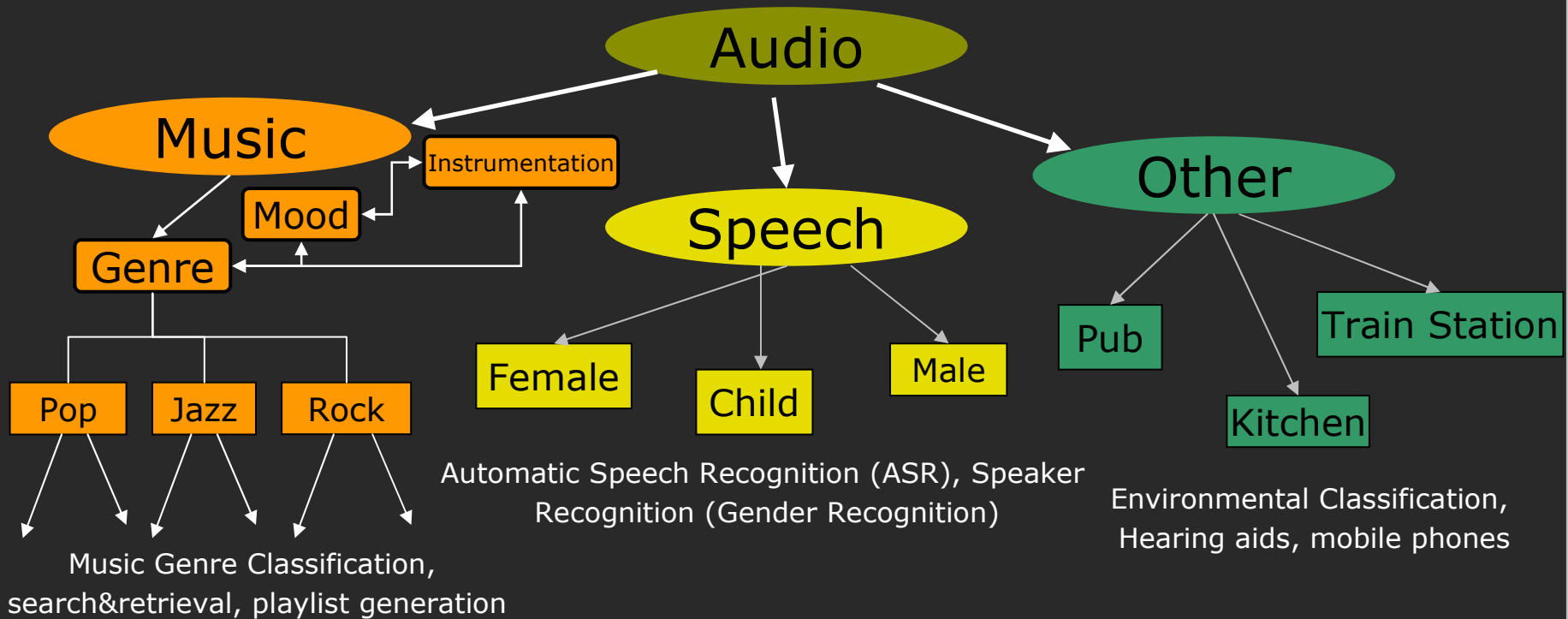
Agenda

- Introduction
- Feature Extraction
- Classification
- Feature integration project [Meng04]
- Summary





Example of Audio Hierarchy





What's the idea ?

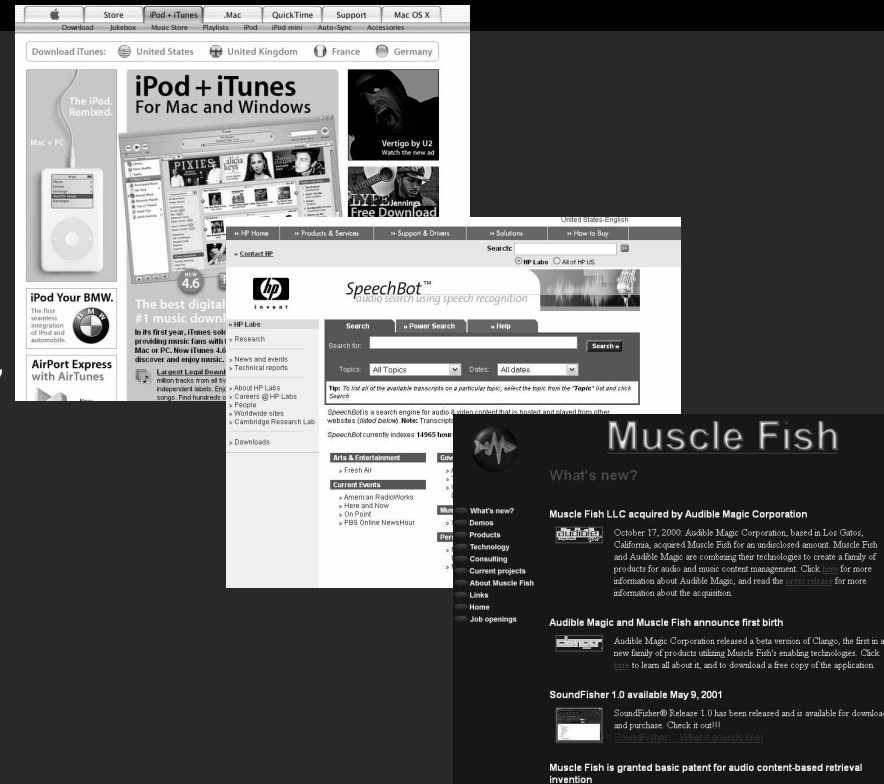
- Music is everywhere – and often described by genre
- Want to be able to automatically annotate music with genre, see e.g. [tzanetakis02]
- Sound features which are interesting in genre classification, may also be interesting in related fields of research – like music search and retrieval, segmentation, playlist generation.





Applications

- Large on-line music stores
 - like Apple's iTunes (1,000,000+ songs), Sony, Amazon, etc.
- Next generations of media management products
 - query-by-humming, search and retrieval, MPEG-7 Format
- Some existing products
 - MuscleFish (now soundfisher at <http://www.soundfisher.com/>)
 - SpeechBot (<http://speechbot.research.compaq.com/>)
 - Findsounds (<http://www.findsounds.com/>).





FindSounds

Search the Web for Sounds

Search for [Help](#)

[Need Examples?](#)

| File Formats | Number of Channels | Minimum Resolution | Minimum Sample Rate | Maximum File Size |
|--|--|--|--|---------------------------------------|
| <input checked="" type="checkbox"/> AIFF | <input checked="" type="checkbox"/> mono | 8-bit <input type="button" value="v"/> | 8000 Hz <input type="button" value="v"/> | 2 MB <input type="button" value="v"/> |
| <input checked="" type="checkbox"/> AU | <input checked="" type="checkbox"/> stereo | | | |
| <input checked="" type="checkbox"/> WAVE | | | | |

Sounds 1-10 of 200 labelled "train"



1. <http://sep800.mine.nu/files/sounds/toytrainhorn.wav>
toy train horn
3k, mono, 8-bit, 8000 Hz, 0.4 seconds ([show page](#) | [e-mail this sound](#))

[Sound Designers, click here](#)



2. <http://www.srim.org/IMAGES/Trainhorn.wav>
train whistle
5k, mono, 8-bit, 11025 Hz, 1.6 seconds ([show page](#) | [e-mail this sound](#))



3. <http://newton.umsl.edu/exhibit/doppler-aux.au>
train whistle
7k, mono, 8-bit, 8000 Hz, 0.8 seconds ([show page](#) | [e-mail this sound](#))

If you like FindSounds.com, you will love FindSounds Palette!



4. <http://www.tux.org/pub/X-Windows/games/freeciv/incoming/sounds/TRAINW.WAV>
train whistle
7k, mono, 8-bit, 11025 Hz, 2.9 seconds ([show page](#) | [e-mail this sound](#))

If you like FindSounds.com, you will love FindSounds Palette!



5. <http://ftp.megamirror.com/pub/games/freeciv/contrib/sounds/dubious/sounds/TRAINW.WAV>
train whistle

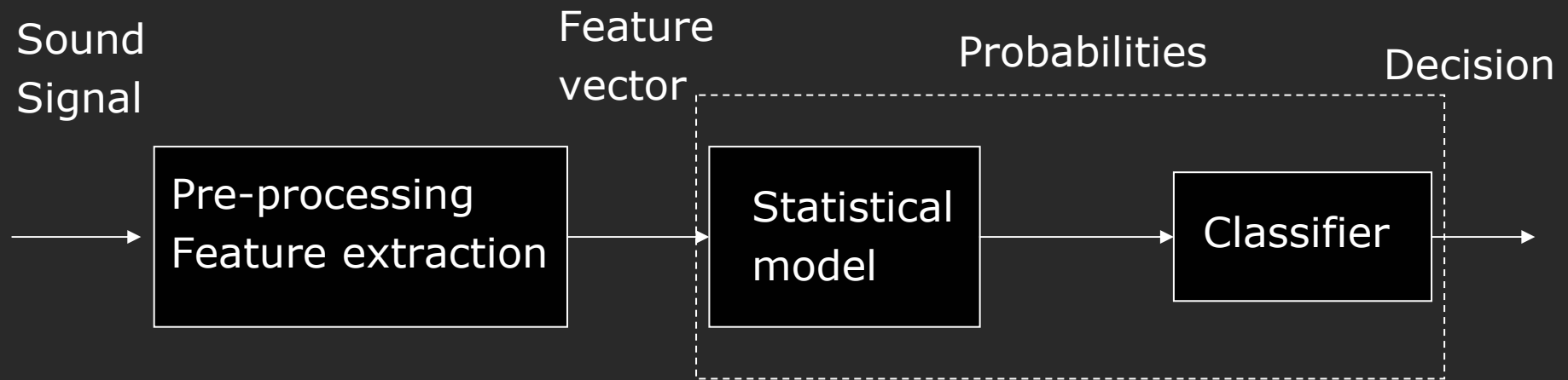


Problems with Music Genre

- Genres are subjectively defined – not an intrinsic property
[Aucouturier02]
 - Different labelling schemes from e.g. Amazon, iTunes, Gracenote, Freedb
 - Cultural background
- Decision-time horizon problems *[Ahrendt04, Meng04]*



Typical Audio Classification System



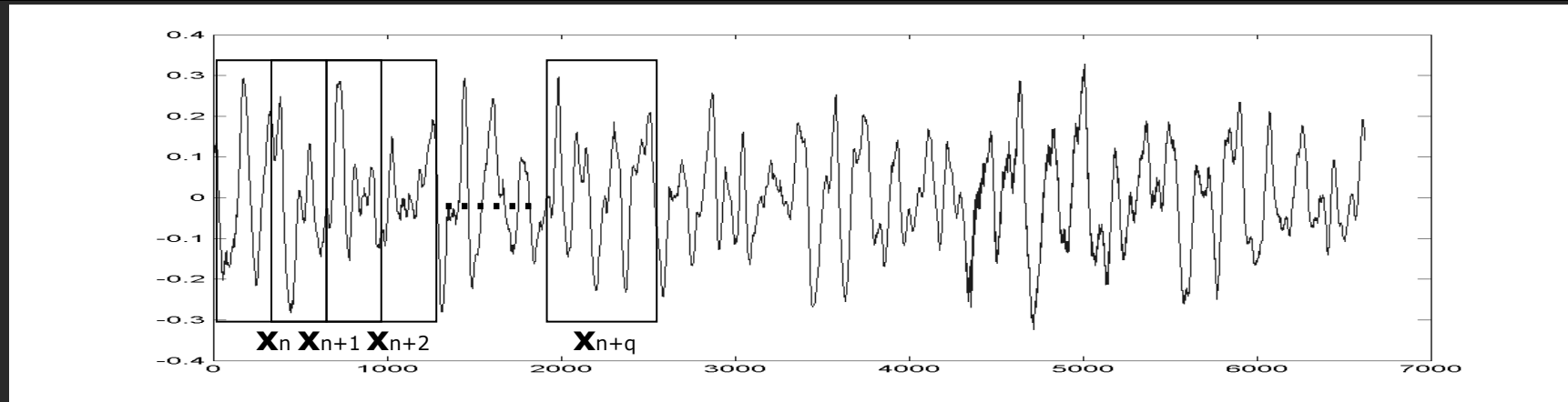


Agenda

- Introduction
- Feature Extraction
- Classification
- Feature integration project *[Meng04]*
- Summary



Feature Extraction



$$X_n = \begin{bmatrix} x_n^1 \\ x_n^2 \\ \vdots \\ x_n^D \end{bmatrix}$$

- Typical frames is 10-30ms long, and normally overlap.
- Time or/and frequency features are extracted. Features at similar time levels are stacked.
- A piece of audio (say p) can be expressed as a $D \times N_p$ matrix.



Some features

■ Short Time Features (10-30 ms)

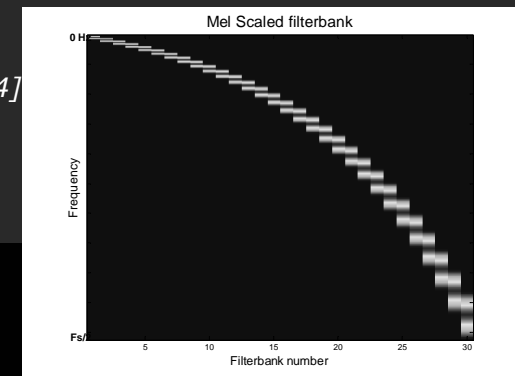
- Mel Frequency Cepstral Coefficients (MFCC, orig. for automatic speech recognition(ASR), see e.g. [Davis80]),
- Linear Predictive Coefficients (LPC, orig. for ASR). See e.g. [Makhoul75]
- Time Zero Crossing Rate (segmentation). See e.g. [Tzanetakis02]
- Short Time Energy (silence detection).
- MPEG-7 features (Audio Spectrum Centroid, Audio Spectrum Spread, Spectral Flatness), see e.g. [Ahrendt04]
- Pitch estimators (Speaker recognition / gender recognition)

■ Sound Texture Windows (400msec. – 1 second)

- Mean/variance (or some other model) of short time features (feature integration), see e.g. [Tzanetakis02]

■ Long time features (several seconds)

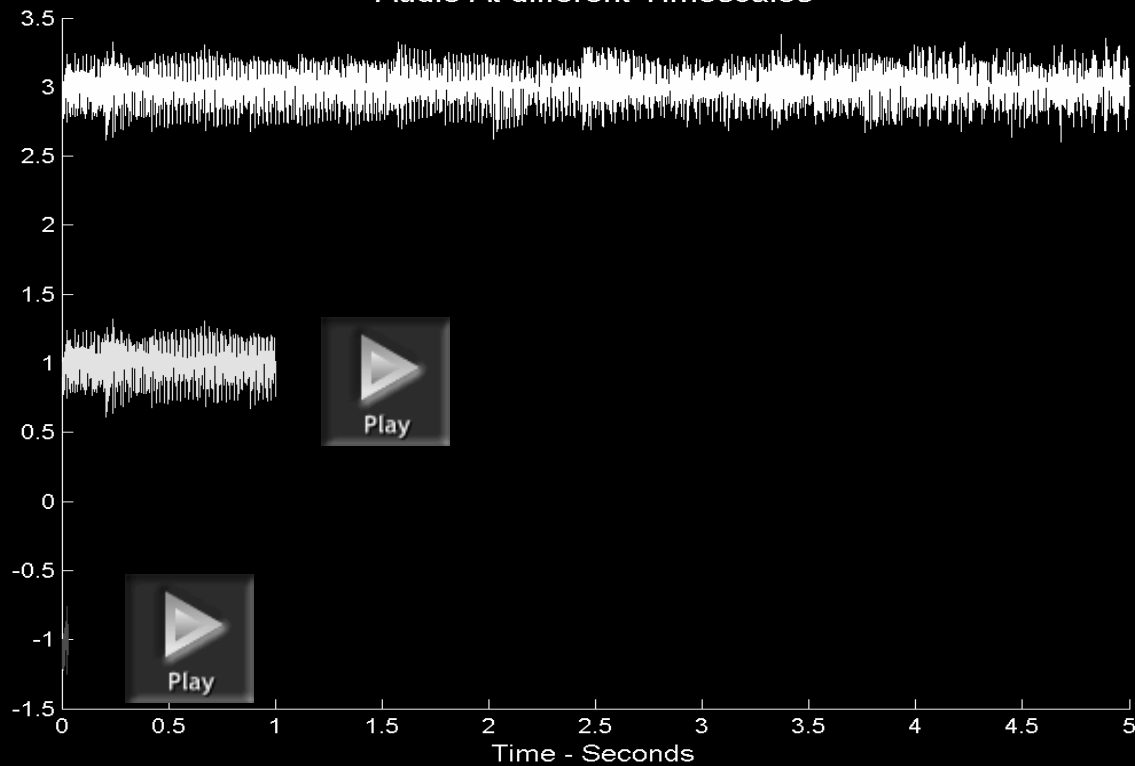
- Beat estimators (Music analysis, transcription) [Tzanetakis02]
- Feature integration from short time features, [Tzanetakis02,Ahrendt04,meng04]





So why does time matter ?

Audio At different Timescales



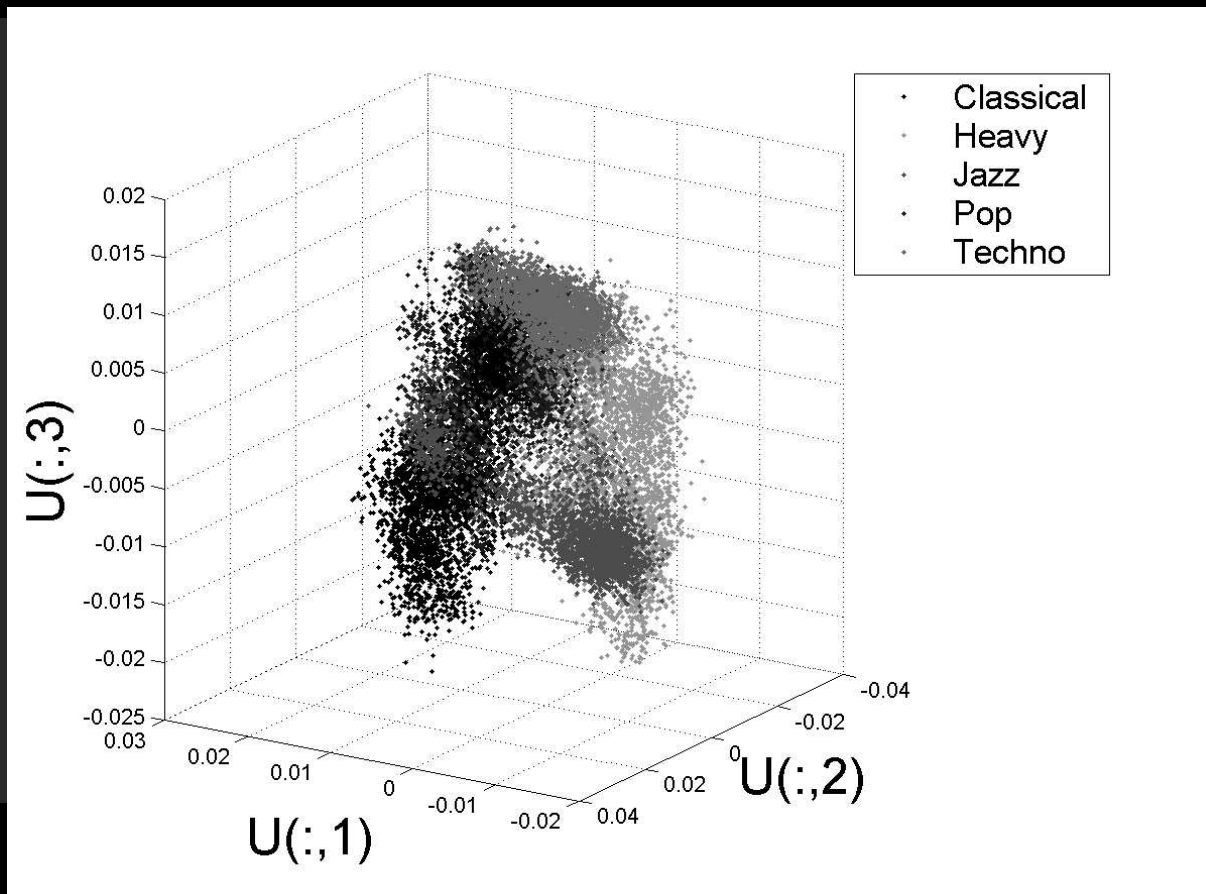


Agenda

- Introduction
- Feature Extraction
- Classification
- Feature integration project [Meng04]
- Summary



Projection of songs



- Each song is 30 sec.



Statistical Model

Desired (posterior probability) : $P(C | \mathbf{x})$ (class C and features \mathbf{x})

Investigated models :

- Linear and non-linear neural networks

- Gaussian distribution

- Gaussian Mixture Model

- Hidden Markov Model



Decision time horizon

- What is early and late information fusion [Ahrendt04]
 - **Early information fusion**: operation on short time features **before** classification (and decision making).
 - **Late information fusion** : assembles information on the basis of the outputs from classifier.

- Early information fusion (Feature integration):
 - Dynamic PCA. [Wu95]
Time stacking of features and perform a PCA as to decorrelate features in both time and between the features.
 - Texture windows
Mean / Variance of features
Model temporal characteristics using e.g. an Autoregressive Model [Meng04]

- Late information fusion [Kittler98] :
 - Sum Rule , Median Rule or Majority voting.



Agenda

- Introduction
- Feature Extraction
- Classification
- Feature integration project [Meng04]
- Summary



Feature Integration project [Meng04], setup

- Goal
 - Music Genre Classification
 - 5 Genres (Rock, Classical, Jazz, Pop, New-Age)
 - Find best features at timescales from 30ms to 5 seconds and compare with human performance

- Features
 - Short time, Mel Frequency Cepstral Coefficients (MFCC)

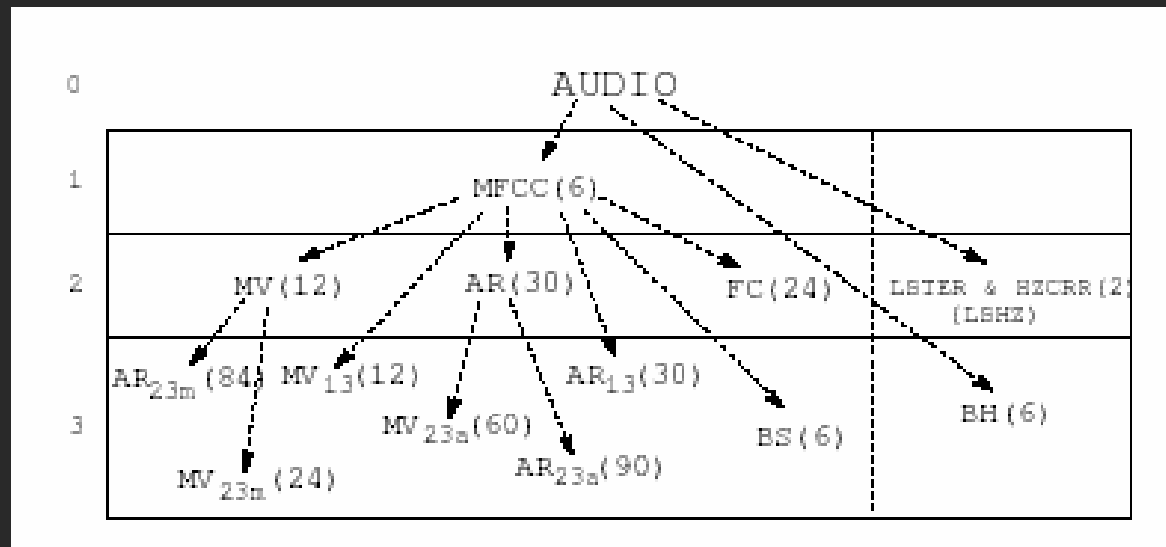
- Method
 - Combining early and late information fusion. Compare with human performance.

- Classifiers
 - Linear Neural Network (Trained Discriminately)
 - No. Parameters : $\sim 5D$
 - Gaussian Classifier (GC)
 - No. Parameters : $\sim 5D(D/2+2)$



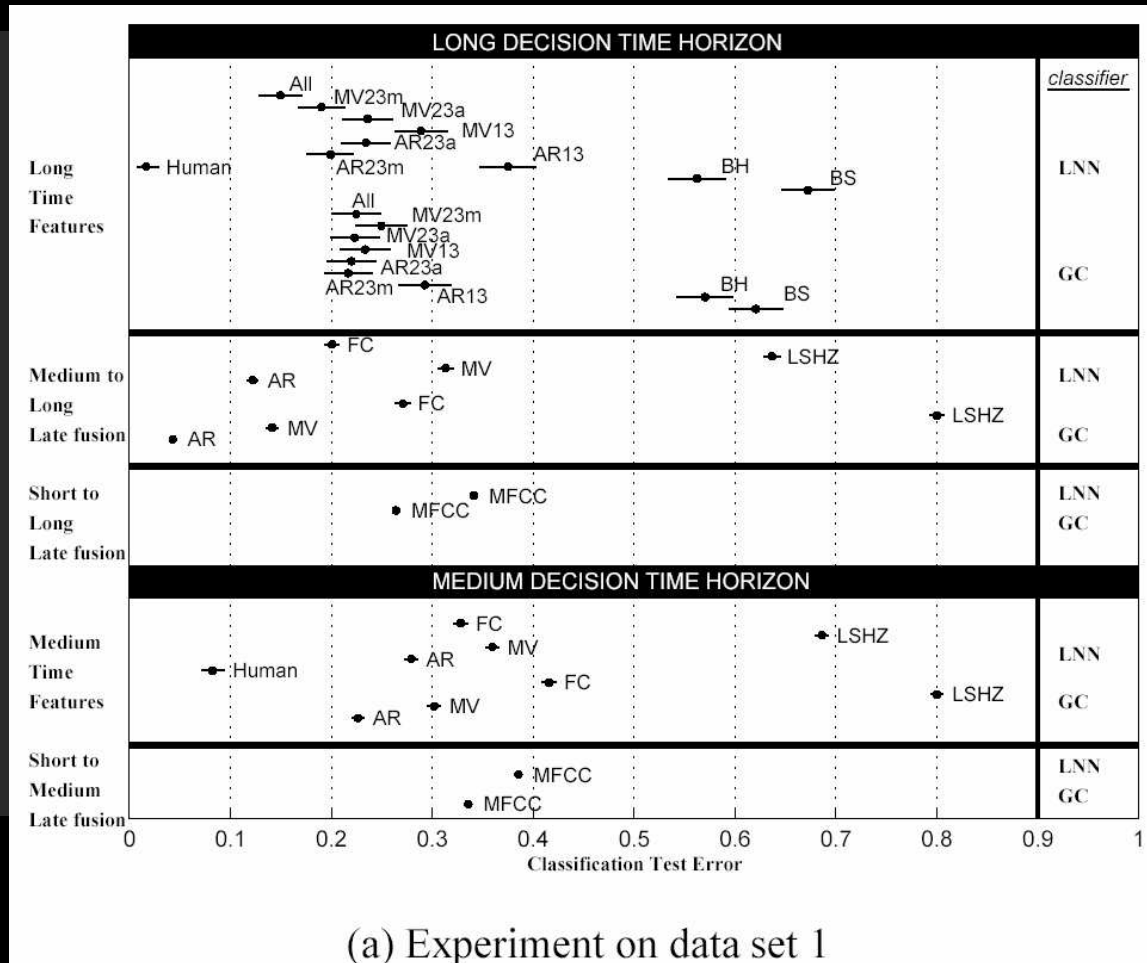
Feature integration project – early fusion

1. ~ 30 ms
2. ~ 1 second
3. ~ 5 seconds





Feature Integration Project, outcome



(a) Experiment on data set 1



Feature Integration Project, outcome

- Combination of feature integration and late information fusion improves performance.
- Temporal information in short time features are important in Music Genre Classification.
- Generalizes to other areas such as playlist-generation, retrieval.



Agenda

- Introduction
- Feature Extraction
- Classification
- Feature integration project
- Summary



Summary

- The area of audio mining is expanding.
- Music Genre Classification can to some extent be used as a test-bench for new music features.
- A more natural approach to music genre is the use of multi-labels.
- Music genre is a first step to music navigation.



References

- Tzanetakis, G. and Cook, P. : "Musical Genre Classification of Audio Signals", *IEEE Transactions on speech and audio processing*, **2002** , 10 , 293-302
- Ahrendt, P. and Meng, A. and Larsen, J. : "Decision Time Horizon for Music Genre Classification using Short-Time Features", *EUSIPCO, Vienna, Austria*, **2004** , 1293-1296
- Davis, S. B. and Mermelstein, P. : "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences", *IEEE Transactions on Acoustics, Speech and Signal Processing*, **1980** , ASSP , 357-366
- Kittler, J. and Hatef, M. and Duin, Robert P.W. and Matas, J. : "On Combining Classifiers", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1998** , 20 , 226-239
- Aucouturier, J-J. and Pachet, F.: "Representing Musical Genre: A State of the Art", *Journal of New Music Research*, **2003** , 32 , 83-93
- W. Ku and R. H. Storer and Georgakis, C.: "Disturbance detection and isolation by dynamic principal component analysis", *Chemometrics and Intelligent Laboratory Systems*, **1995** , 30 , 179-196
- Meng, A and Ahrendt, P. and Larsen, J. : " Improving Music Genre Classification by Short-Time Feature Integration", Submitted to ICASSP 2005, Philadelphia, USA.