Musical Subgenre Classification

Topics in Computer Music
Summer semester 2016

Ali Tarzan
ali.tarzan@rwth-aachen.de

Overview

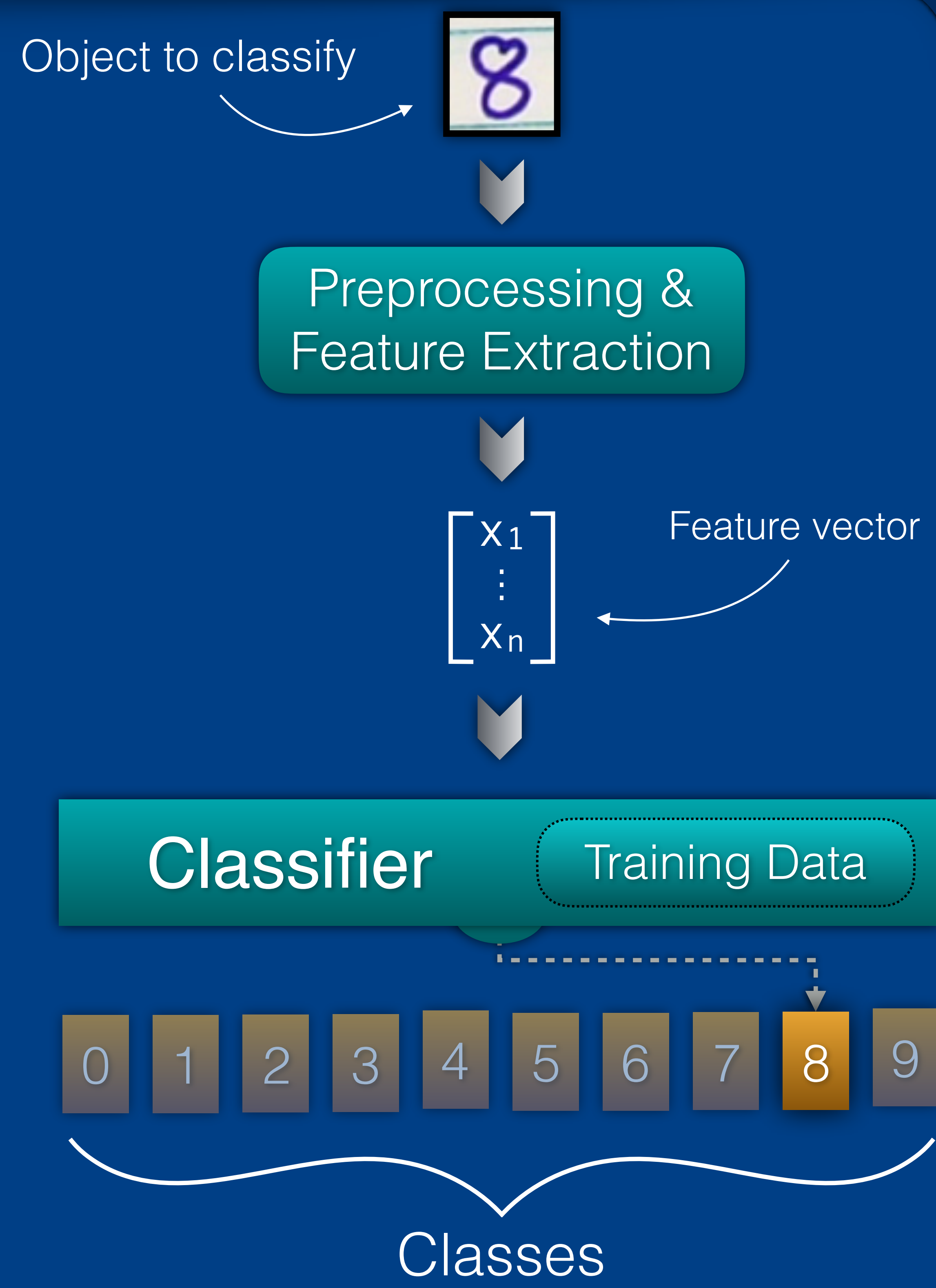
- Classifiers in general
- Musical Genre Classification
- Problems with classifying musical genres
- Approaches to genre classification
- Demo

Introduction: Classification in general

- **Input:**
  n-dimensional vector x generated from
  - pixels of an image
  - samples of `.wav` file of a music track
  - financial data of a person
  - …

- **Output:**

  Which class does x belong to?

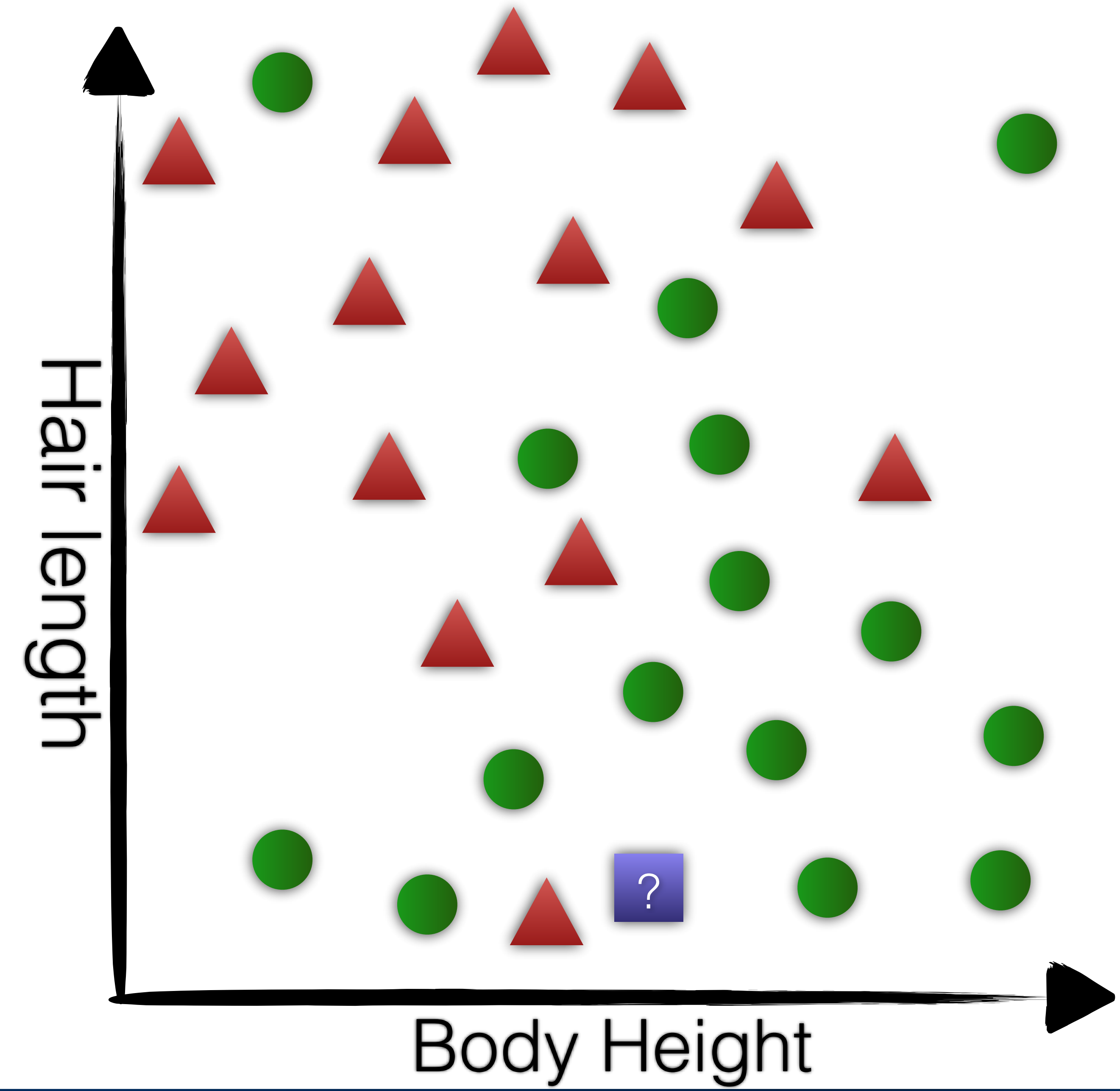


Heavily researched area with numerous algorithms!

Features and training data example

Classes:
- : Female
- : Male

Features:
- [x] Hair length
- [x] Body Height
- [ ] ~~Age~~ *(useless)*
- [ ] ~~Hair Color~~ *(useless)*

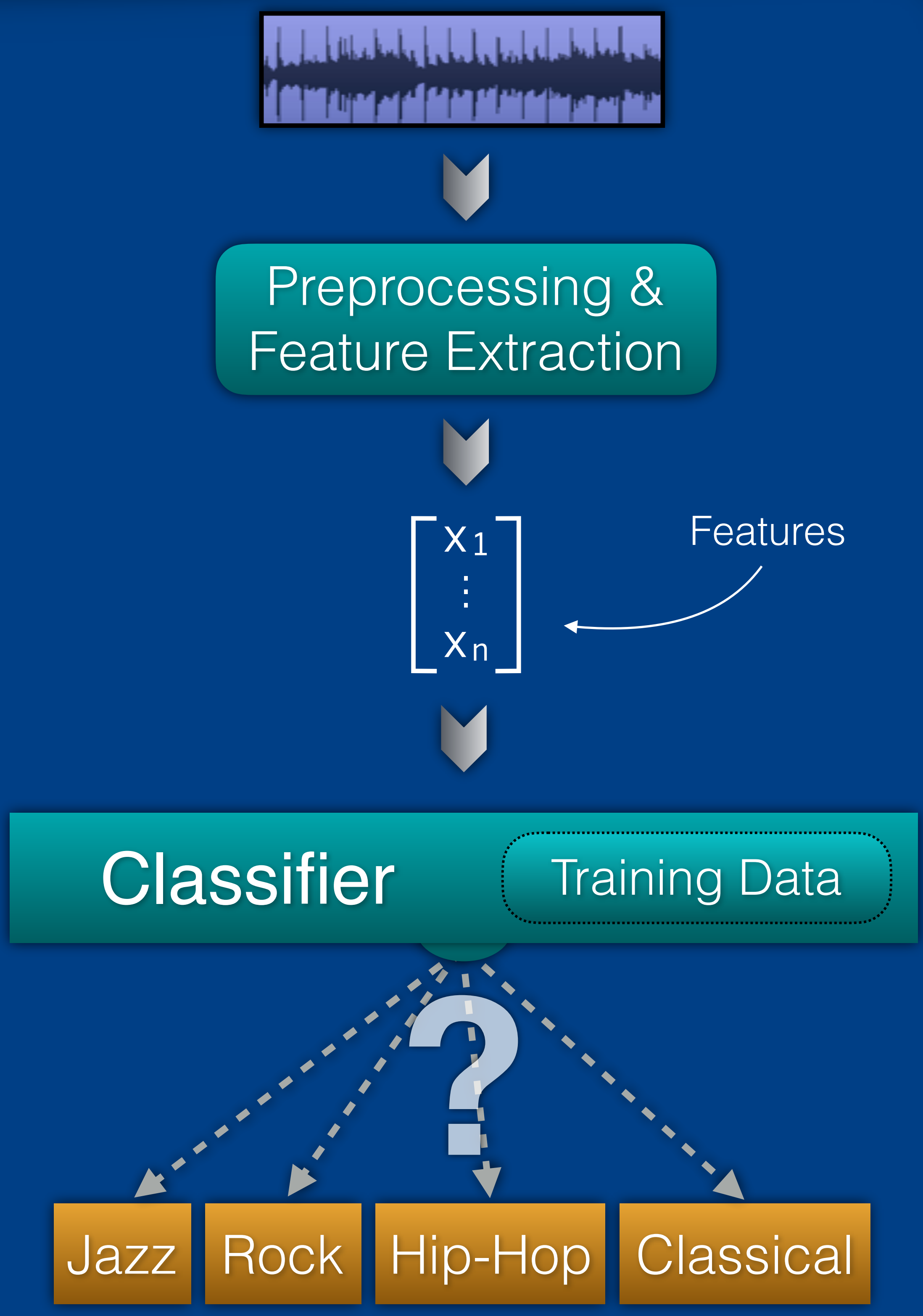


Goal: Learn mapping from feature space to classes from training data

Musical Genre Classification

- Which musical genre does input signal belong to?

- Example Features:
    - Loudness
    - Bandwidth
    - Zero crossing rate
    - Pitch histogram based features
    - Entropy



Musical Genre Classification: Overview

- First occasional works since the 1990s
- **1995, Matityaho**:
  100% success rate, but only Classical and Pop as genres.
- Subject became more popular in the 2000s …
- … but stopped making notable progress a few years later:

> “ Automatic genre classification performance appears to have fallen into a local maximum recently, and serious modifications to the approaches used are needed in order to realize further improvements. ”

Musical genre classification: Is it worth pursuing and how can it be improved?
- Cory McKay, Ichiro Fujinaga (2006)

- **Today**: Still ongoing research, problem not satisfyingly solved yet

Problems with the term “Genre”

- Artists do not intend their songs to belong to a specific genre ⇒ Which is correct genre?
- Genres are not always disjoint
- Some genres naturally closer to each other than other genres

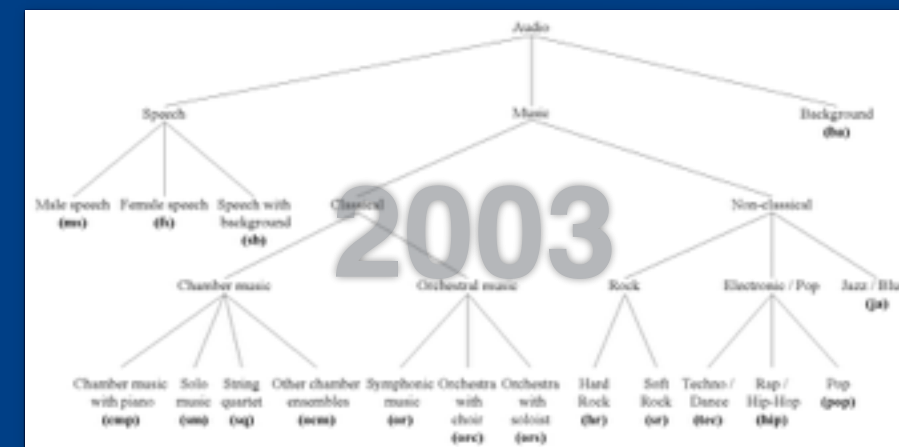
Comparability between results

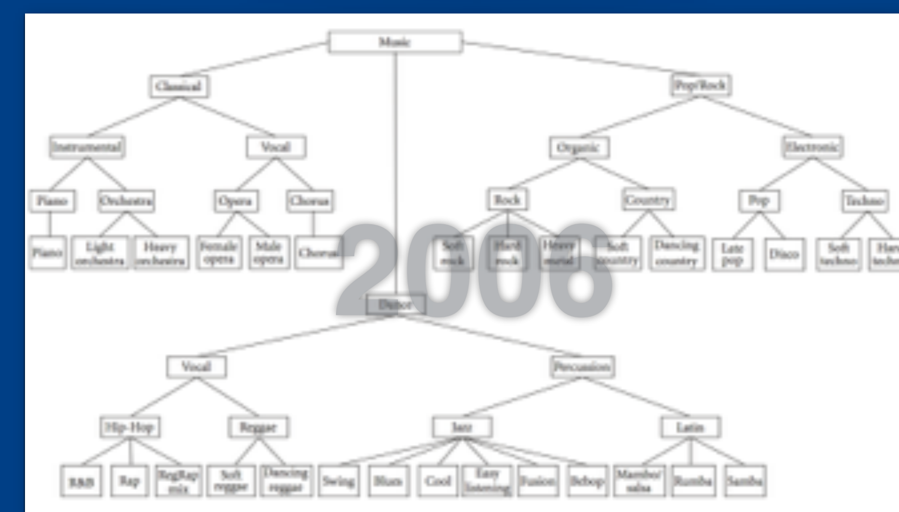
- No general consensus on genre taxonomy
- Studies employ their own taxonomy
- Recent studies on special techniques use rather simple taxonomies

⇒ Results not comparable

⇒ Progress in the area hard to measure

⇒ State of the art hard to determine

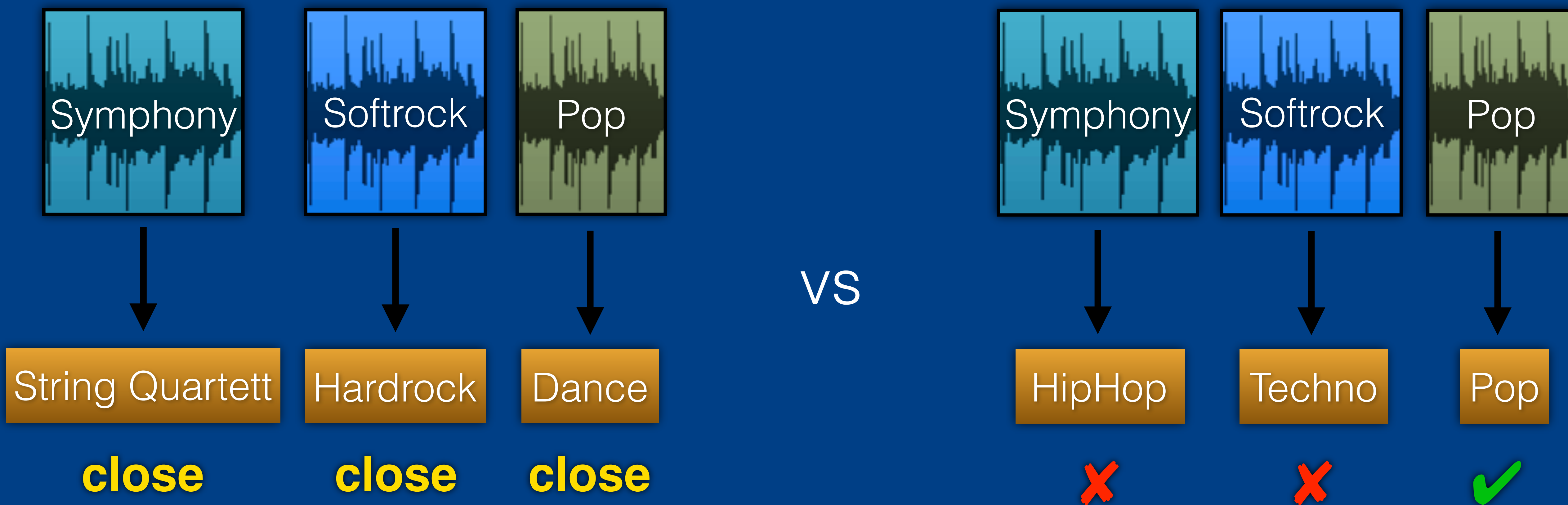




2015

Genre	*Bhangra*	*Ghazal*	*Classical*
Bhangra	*66*	0	0
Ghazal	0	*70*	0
Classical	0	0	*70*
Folk	2	3	1
Jazz	0	1	1
Pop	4	3	0
Rock	0	2	0

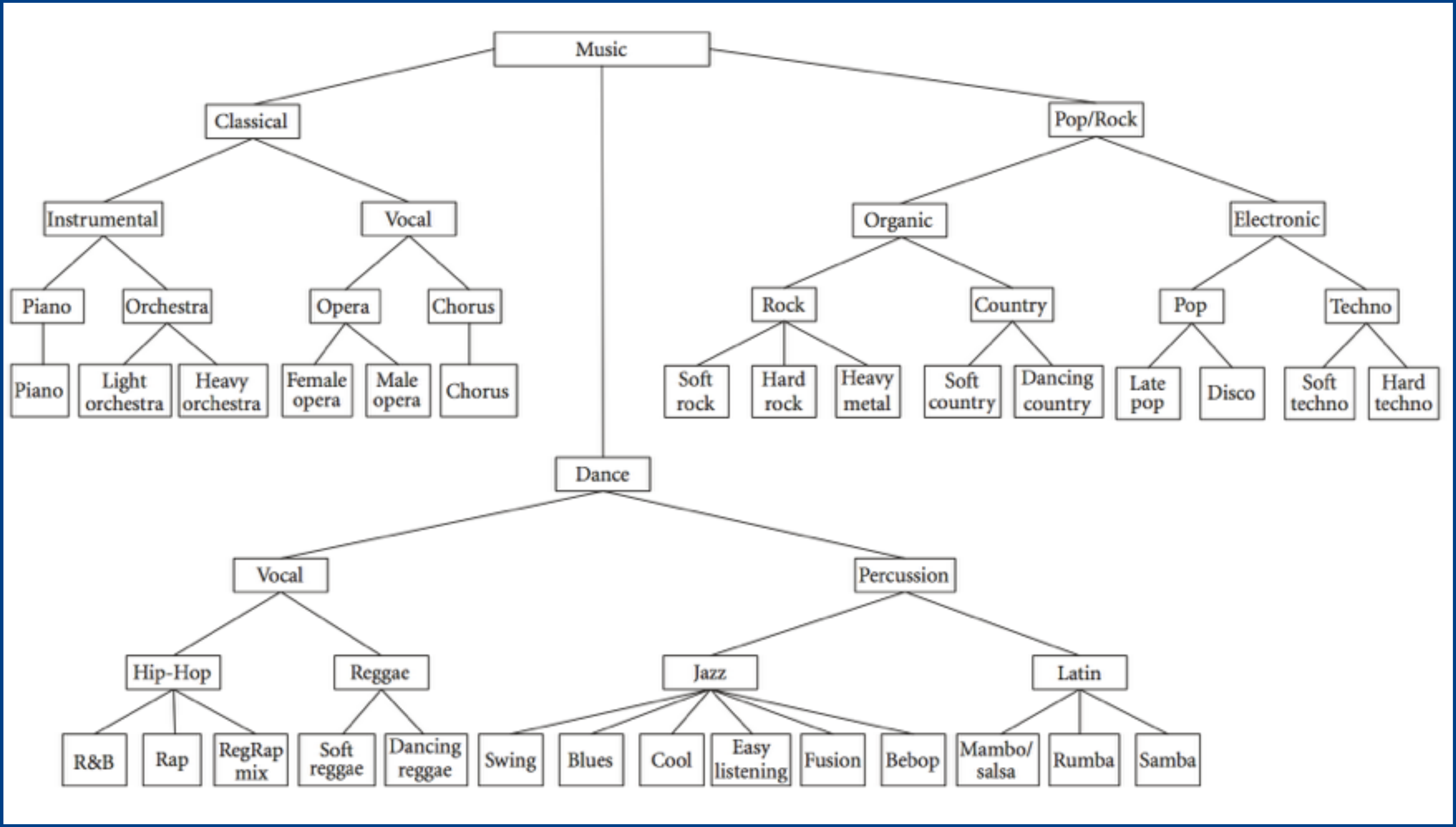
Classification Quality



Error rate alone not a good measure for classification quality!

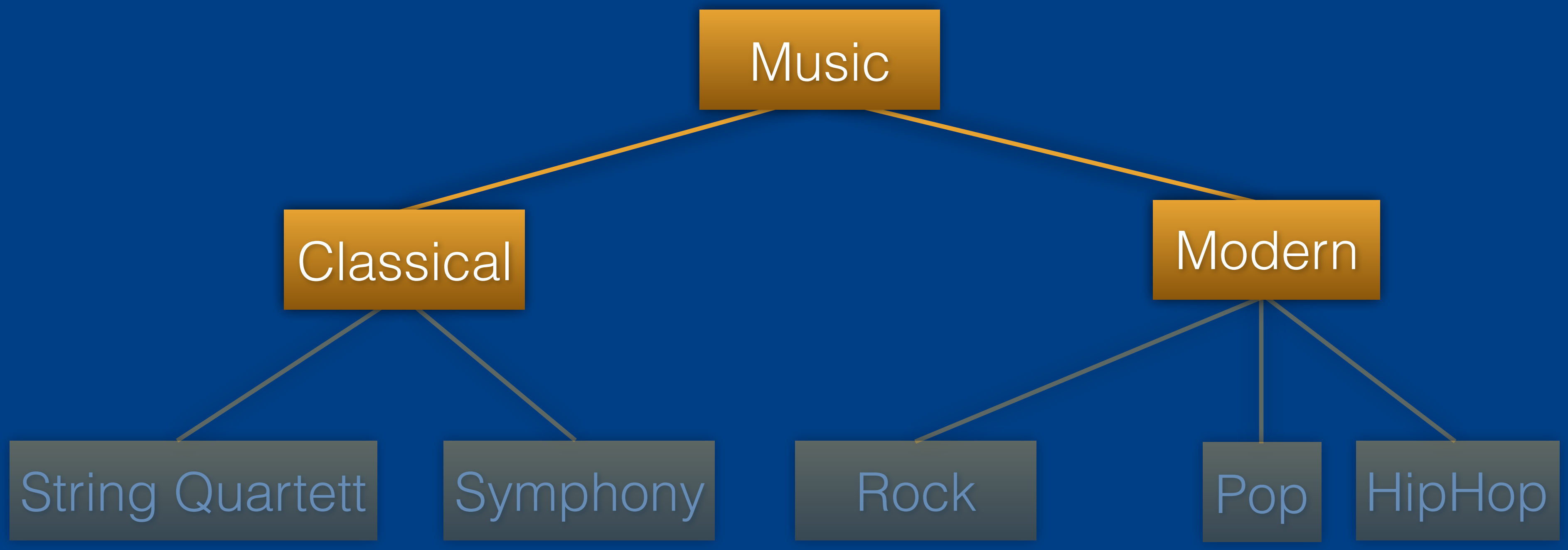
Do not ignore how acceptable an error ist

Example hierarchy of genres



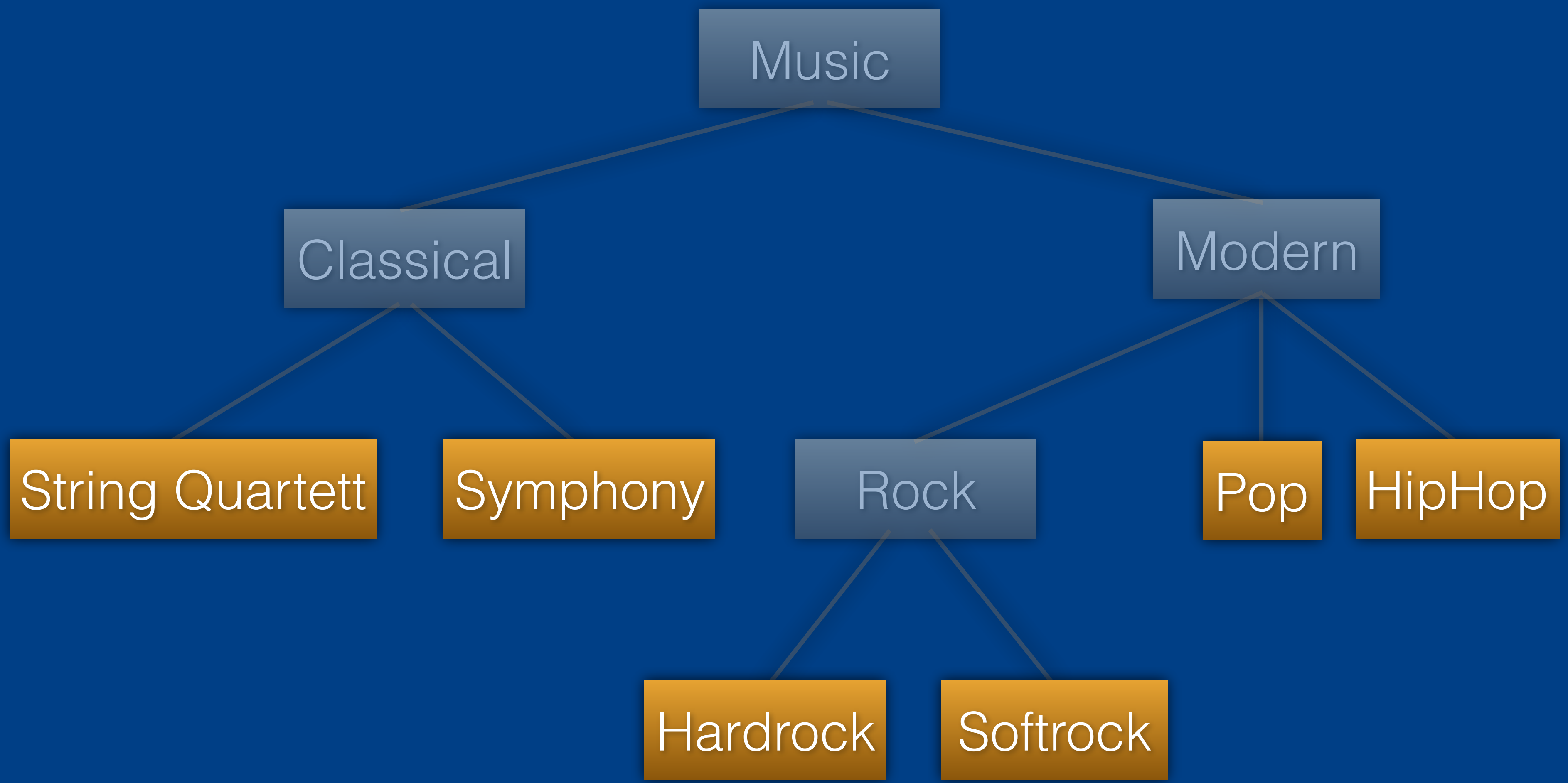
Top-down approach (Burred & Lerch 2003)

- Start at highest layer
- Next layer with subgenre-specific subset of features
- Smaller set of features at once:
  - Different features more relevant for different subgenres
  - Counters problem of dimensionality



Bottom-up approach
(Barbedo & Lopes 2006)

- Consider only leaf classes
- Higher levels implicitly classified, e.g. Softrock ⇒ Rock



Barbedo & Lopes: Feature Extraction

Frames of 21.3 ms (= 1024 samples)

Extract 4 features for each frame

$$\begin{bmatrix}a\\b\\c\\d\end{bmatrix}\ \begin{bmatrix}a\\b\\c\\d\end{bmatrix}\ \begin{bmatrix}a\\b\\c\\d\end{bmatrix}\ \begin{bmatrix}a\\b\\c\\d\end{bmatrix}\ \begin{bmatrix}a\\b\\c\\d\end{bmatrix}\ \begin{bmatrix}a\\b\\c\\d\end{bmatrix}\ \begin{bmatrix}a\\b\\c\\d\end{bmatrix}\ \begin{bmatrix}a\\b\\c\\d\end{bmatrix}\ \ldots$$

Groups of 1 second segments

$$\left\{\begin{bmatrix}a\\b\\c\\d\end{bmatrix}, \begin{bmatrix}a\\b\\c\\d\end{bmatrix}, \ldots\right\}, \left\{\begin{bmatrix}a\\b\\c\\d\end{bmatrix}, \begin{bmatrix}a\\b\\c\\d\end{bmatrix}, \ldots\right\}, \ldots$$

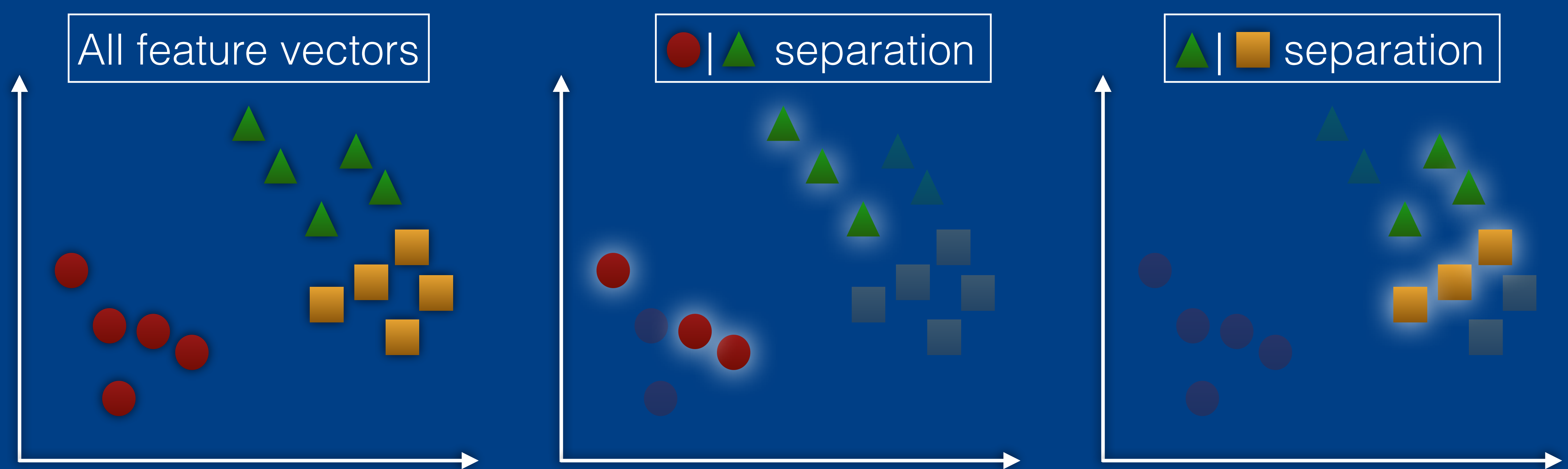
3 summary features for each group

$$\begin{bmatrix}a_1\\b_1\\c_1\\d_1\\\vdots\\d_3\end{bmatrix}, \quad \begin{bmatrix}a_1\\b_1\\c_1\\d_1\\\vdots\\d_3\end{bmatrix}, \ldots$$

Result: Vector of length 12 for every second in signal

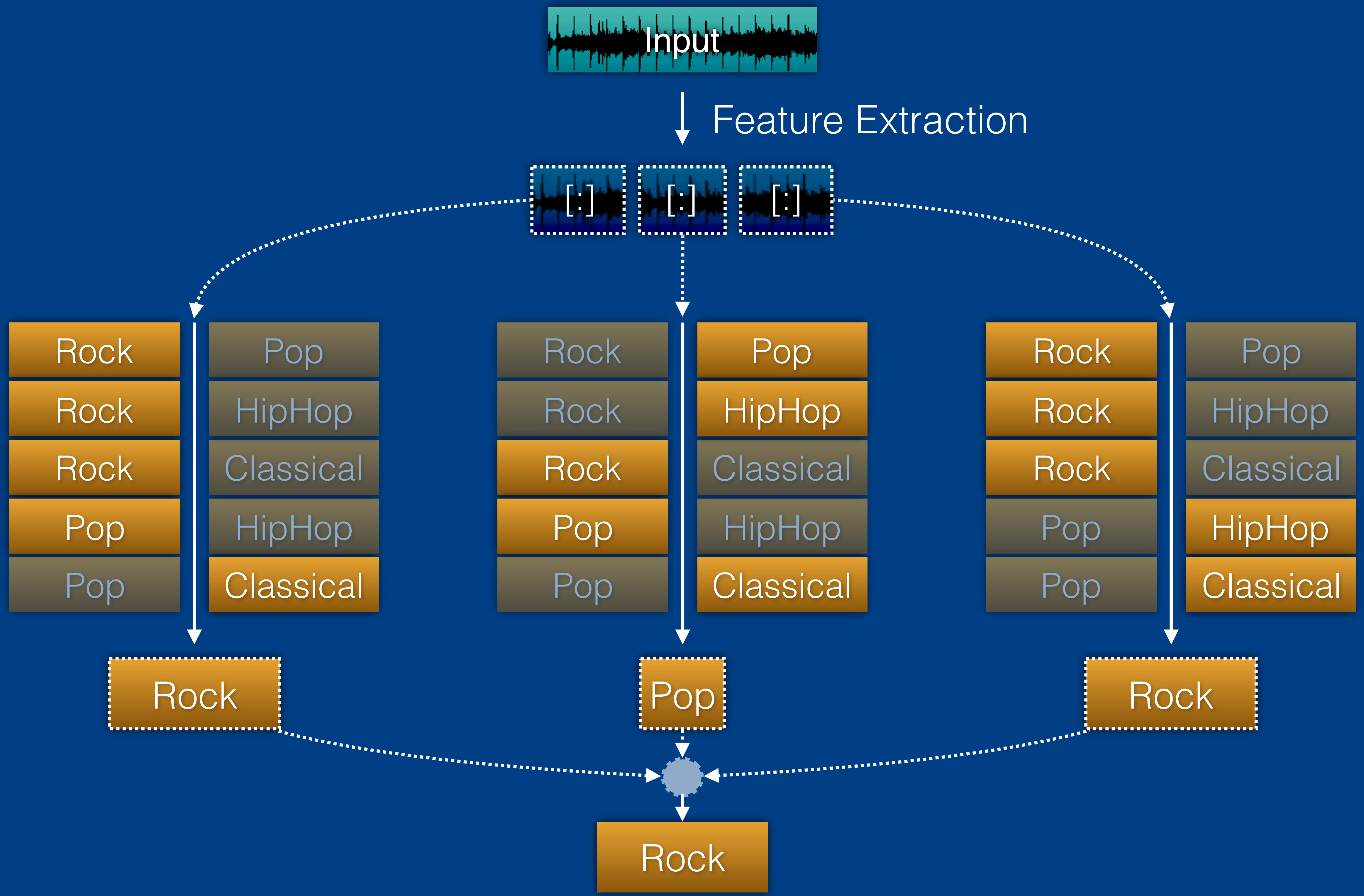
Barbedo & Lopes: Decision Process

- Pairwise decision between classes for each input feature vector
- Selected 3+3 reference vectors **for each pair** of genres in training (by brute forcing best combination out of handpicked signals)
- Class with nearest reference vector wins (euclidian distance)

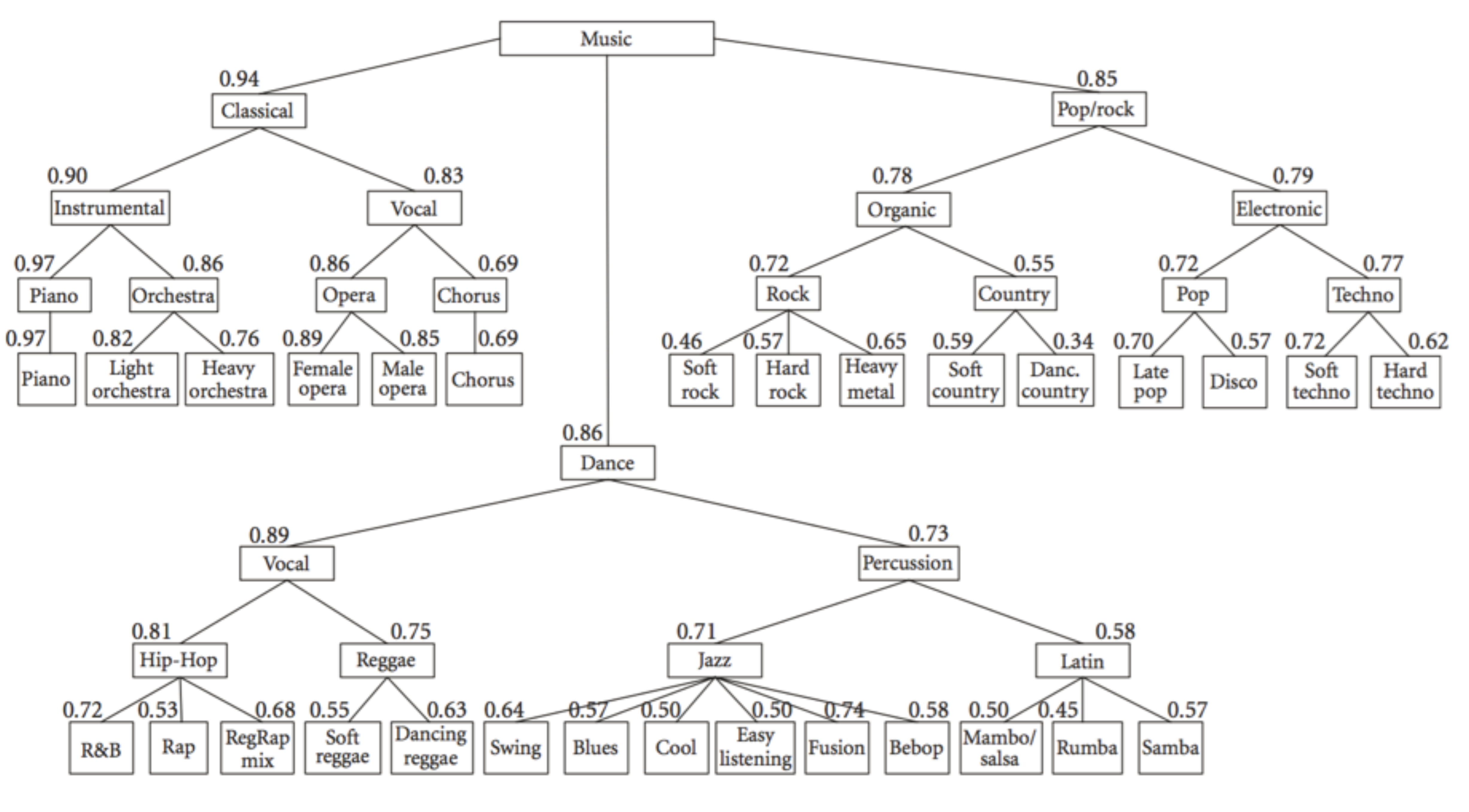


Good results for separating between close subgenres!

Barbedo & Lopes: Decision Process (2)

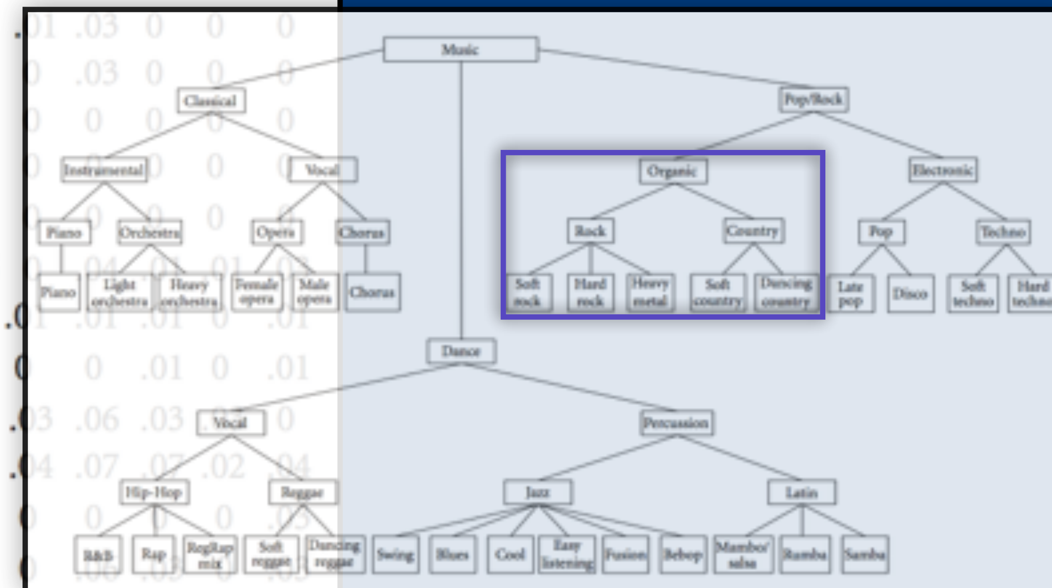


Results (Accuracy)



Results (Confusion Matrix)

	LO	HO	PI	FO	MO	CH	SR	RO	HM	SC	DC	PO	DI	ST	HT	RB	RA	RR	RE	DR	SW	BL	CO	EL	FU	BE	MA	RU	SA
LO	.82	.08	.05	.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.02	0	0	0	0	0	0	0	0
HO	.07	.76	0	0	.02	.04	0	0	0	.02	0	0	0	0	0	0	0	0	0	0	.02	0	.03	0	.01	.03	0	0	0
PI	0	0	.97	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.03	0	0	0
FO	0	.06	0	.89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.03	0	.02	0	0	0	0	0	0
MO	0	.07	0	0	.85	.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.02	.02	.02	0	0	[illegible]	0	0	0
CH	.06	.19	0	0	.03	.69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.03	0	0	0	0	[illegible]	[illegible]	0	[illegible]
SR	0	.01	0	0	0	0	.46	.08	0	.12	.04	.06	0	0	0	.03	.01	0	.01	.01	0	.01	.01	.06	0	.04	[illegible]	.01	[illegible]
RO	0	.01	0	0	0	0	.05	.57	.13	.01	.04	.05	.01	.01	.02	0	0	0	0	.02	0	.04	0	0	.01	.01	.01	0	.01
HM	0	0	0	0	0	0	.01	.25	.65	0	.02	.01	0	0	.02	0	0	0	0	0	0	.02	0	0	0	0	.01	0	.01
SC	0	.01	0	0	0	0	.07	.01	0	.59	.11	0	0	0	0	0	0	0	0	0	0	0	.01	.05	.03	.06	.03	[illegible]	0
DC	0	.01	0	0	.02	0	.09	.15	0	.04	.34	.05	0	0	0	.02	0	0	0	.02	0	.01	0	.01	.04	.07	.07	.02	.04
PO	0	0	0	0	0	0	.04	.08	0	0	.03	.70	.07	.03	0	0	.01	0	0	.01	0	0	0	0	0	0	[illegible]	0	[illegible]
DI	0	0	0	0	0	0	0	.11	0	0	.06	.06	.57	0	0	0	.03	0	0	.02	0	.03	0	0	0	[illegible]	[illegible]	[illegible]	[illegible]
ST	0	0	0	0	0	0	0	.04	0	0	0	.06	.02	.72	.10	0	0	0	.02	.04	0	0	0	0	0	0	0	0	0
HT	0	0	0	0	0	0	0	.01	.05	.02	0	.03	0	.13	.62	.01	.05	.01	0	.07	0	0	0	0	0	0	0	0	0
RB	0	0	0	0	0	0	.03	0	0	0	0	.03	0	0	0	.72	.03	.04	.03	.09	0	0	0	0	0	0	0	0	.03
RA	0	0	0	0	0	0	0	0	0	0	0	.01	0	.02	.02	.09	.53	.19	0	.12	0	0	0	0	0	0	.02	0	0
RR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.06	.10	.68	.03	0	0	0	0	0	0	0	.10	0	.03
RE	0	0	0	0	0	0	0	0	0	0	0	.03	0	0	0	0	0	.06	.55	.27	0	0	0	0	0	.09	0	0	0
DR	0	0	0	0	0	0	0	0	0	0	0	0	.04	.02	.02	.04	0	.09	.08	.63	0	0	0	0	0	.06	.02	0	0
SW	0	.03	0	0	.04	0	.03	0	0	0	.03	.03	0	0	0	0	0	0	0	0	.64	0	.03	0	.07	.03	0	0	.07
BL	0	0	0	0	0	0	.05	.13	0	.03	0	.03	0	0	0	0	0	0	0	.05	0	.57	0	.08	.03	0	0	0	.03
CO	0	.05	.03	0	0	0	.05	0	0	.08	0	0	.05	0	0	0	0	0	.03	.03	0	0	.50	.10	0	.08	0	0	0
EL	0	.03	0	0	0	0	.08	.03	0	.10	0	0	.03	0	0	0	0	0	0	0	.02	0	.11	.50	0	.05	0	.02	.03
FU	0	0	0	0	0	0	.03	.05	0	0	.02	.05	.03	0	0	0	0	0	0	0	0	0	.03	0	.74	.05	0	0	0
BE	0	.06	0	0	0	0	.03	0	0	.03	.03	0	0	0	0	0	.03	0	.03	.03	0	.03	0	.03	0	.58	.09	.03	0
MA	0	0	0	0	0	0	0	0	0	0	.07	0	.01	.01	0	.03	.03	.07	0	.04	0	.01	0	.01	.01	.12	.50	0	.09
RU	0	.02	0	0	0	0	.10	.02	0	.03	.02	.02	0	0	0	.02	0	0	0	.05	0	0	0	.05	.03	.12	0	.45	.07
SA	0	0	0	0	.02	0	.03	.04	0	0	.04	.08	.04	0	0	0	.02	.04	0	.04	0	0	0	0	0	.04	.04	0	.57



Errors mostly close to diagonal (⇒ within same superclass)

Demo: Realtime Genre Classifier

Source

MARSYAS Genre Meter
by George Tzanetakis
http://marsyas.info/

George Tzanetakis, Georg Essl, Perry Cook:

Automatic Musical Genre Classification Of Audio Signals

Recap, Remarks & Outlook

- Music tracks can not always be assigned to exactly one class
- Classes/genres are not necessarily disjunct
- Genres can have a hierarchical structure

- Advantages of hierarchical genre taxonomies:
  Focus on classification quality instead of just error rate

- Possible improvements:
  - Assign several classes to a song with weights
  - Classify different parts of a song independently

Thank you for listening

Questions?

References

- **Automatic Genre Classification of Musical Signals**
  - Jayme Garcia, Arnal Barbedo, Amauri Lopes (2006)
- **Automatic Musical Genre Classification Of Audio Signals**
  - George Tzanetakis, Georg Essl, Perry Cook (2001)
- **Neural Network based model for Classification of Music Type**
  - Benyamin Matityaho (1995)
- **A hierarchical approach to Automatic Musical genre Classification**
  - Juan José Burred, Alexander Lerch (2003)
- **Automatic genre classification of music content: a survey**
  - Nicolas Scaringella (2006)
- **Automatic Subgenre Classification of Heavy Metal Music**
  - Valeri Tsatsishvili (2011)
- **Perceptual feature-based song genre classification using RANSAC**
  - Arijit Ghosal, Rudrasis Chakraborty, Bibhas Chandra Dhara (2015)

Additional Slides

6. CLASSIFICATION STRATEGY

The features extracted for each frame are grouped according to 1-second analysis segments. Therefore, each group will have 92 elements, from which three summary features are extracted: mean, variance, and main peak prevalence which is calculated according to

$$p_{ft}(j) = \frac{\max\left[ft(i,j)\right]}{(1/I)\cdot\sum_{i=1}^{I} ft(i,j)}, \tag{8}$$

where $ft(i,j)$ corresponds to the value of feature ft in the frame i of segment j, and I is the number of frames into a segment. This summary feature aims to infer the behavior of extreme peaks with relation to the mean values of the feature. High p_{ft} indicate the presence of sharp and dominant peaks, while small p_{ft} often means a smooth behavior of the feature and no presence of high peaks.

As a result of this procedure, each segment will lead to 12 summary features, which are arranged into a test vector to be compared to a set of reference vectors. The determination of the reference vectors is described next.

Features used by Barbedo & Lopes

5.1. Spectral roll-off

This feature determines the frequency R_i for which the sum of the spectral line magnitudes is equal to 95% of the total sum of magnitudes, as expressed by [22]:

$$\sum_{k=1}^{R_i} |X_i(k)| = 0.95 \cdot \sum_{k=1}^{K} |X_i(k)|, \tag{1}$$

where $|X_i(k)|$ is the magnitude of spectral line k resulting from a (discrete Fourier transform DFT) with 1024 samples applied to the frame i and K is half the total number of spectral lines (second half is redundant).

5.2. Loudness

The first step to calculate this feature is modeling the frequency response of human outer and middle ears. Such a response is given by [23]

$$\begin{aligned} W(k) = {} & -0.6 \cdot 3.64 \cdot f(k)^{-0.8} - 6.5 \cdot e^{-0.6\cdot(f(k)-3.3)^2} \\ & + 10^{-3} \cdot f(k)^{3.6}, \end{aligned} \tag{2}$$

where $f(k)$ is the frequency in kHz given by

$$f(k) = k \cdot d, \tag{3}$$

and d is the difference in kHz between two consecutive spectral lines (in the case of this work, 46.875).

The frequency response is used as a weighting function that emphasizes or attenuates spectral components according to the hearing behavior. The loudness of a frame is calculated according to

$$\mathrm{ld}_i = \sum_{k=1}^{K} |X_i(k)|^2 \cdot 10^{W(k)/20}. \tag{4}$$

5.3. Bandwidth

This feature determines the frequency bandwidth of the signal, and is given by [19]

$$\mathrm{bw}_i = \sqrt{\frac{\sum_{k=1}^{K} \left[(\mathrm{ce}_i - k)^2 \cdot |X_i(k)|^2 \right]}{\sum_{k=1}^{K} |X_i(k)|^2}}, \tag{5}$$

where ce_i is the spectral centroid of frame i, given by

$$\mathrm{ce}_i = \frac{\sum_{k=1}^{K} k \cdot |X_i(k)|^2}{\sum_{k=1}^{K} |X_i(k)|^2}. \tag{6}$$

Equation (5) gives the bandwidth in terms of spectral lines. To get the value in Hz, bw must be multiplied by d.

5.4. Spectral Flux

This feature is defined as the quadratic difference between the logarithms of the magnitude spectra of consecutive analysis frames and is given by [1]

$$fe_i = \sum_{k=1}^{K} \left\{ \log_{10} [X_i(k)] - \log_{10} [X_{i-1}(k)] \right\}^2. \tag{7}$$