PHILIPS

Features for Audio and Music Classification

Martin F. McKinney and Jeroen Breebaart

Auditory and Multisensory Perception, Digital Signal Processing Group

Philips Research Laboratories

Eindhoven, The Netherlands

PHILIPS

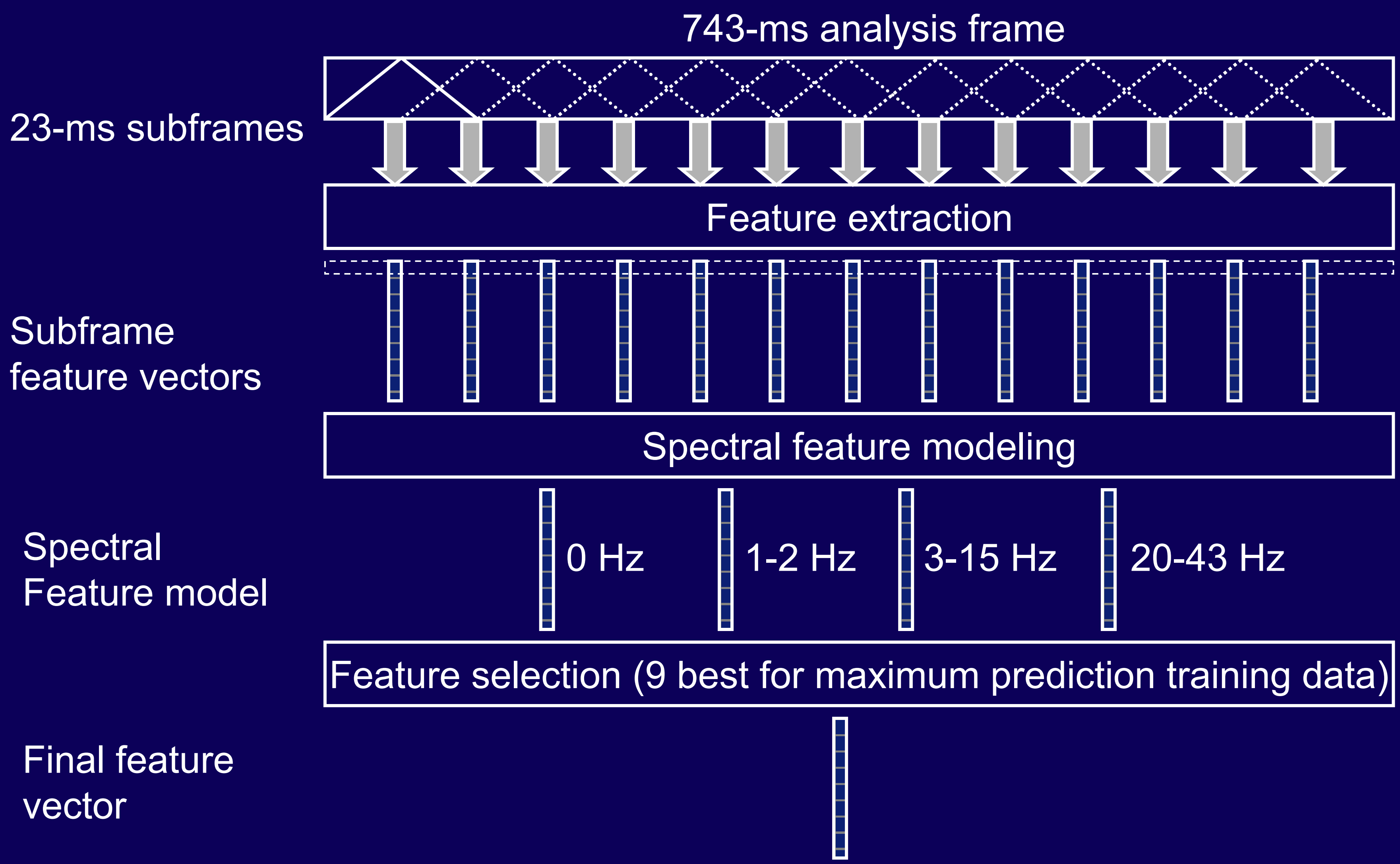
Introduction

- Wanted: automatic audio and music classifier
- Previous work:
  - Typical method: Feature extraction followed by classification
  - Specific method of classification is not always crucial
    - i.e., features are the limiting factor
  - Temporal properties of audio are important for classification and summarization
- Our focus here is on *features* for audio classification and their temporal properties

Method: General

- Compare classification performance of four feature sets:
  - “Standard” low-level signal parameters
  - Mel-frequency cepstral coefficients (MFCC)
  - Psychoacoustic features
  - Auditory filterbank temporal envelope
- Include statistics of feature temporal behavior as additional features
- Evaluate classification using a multivariate Gaussian framework (Quadratic Discriminate Analysis - QDA)

Method: Feature extraction



PHILIPS

Method: Classification

- Classification tasks
  - Five class general audio classification
    - Classical music (35), popular music (188), speech (31), background noise (25), crowd noise (31)
  - Seven class music genre classification
    - Jazz (38), Folk (23), Electronica (27), R&B (43), Rock (37), Reggae (11), Vocal (9)
- QDA training and cross-validation with the *.632+ bootstrap* method

PHILIPS

Results: Standard Low Level features

Feature ranking: General Audio, Music Genre

	DC	1-2 Hz	3-15 Hz	20-43 Hz
1. RMS level	3, 3		8	7, 9
2. Spectral centroid				
3. Bandwidth	6, 7			
4. Zero crossing rate	4			
5. Spectral roll-off freq	1, 2			
6. Band energy ratio	2, 6		4, 1	
7. Delta spectrum mag.				
8. “Pitch”	5, 5		8	
9. “Pitch” strength	9			

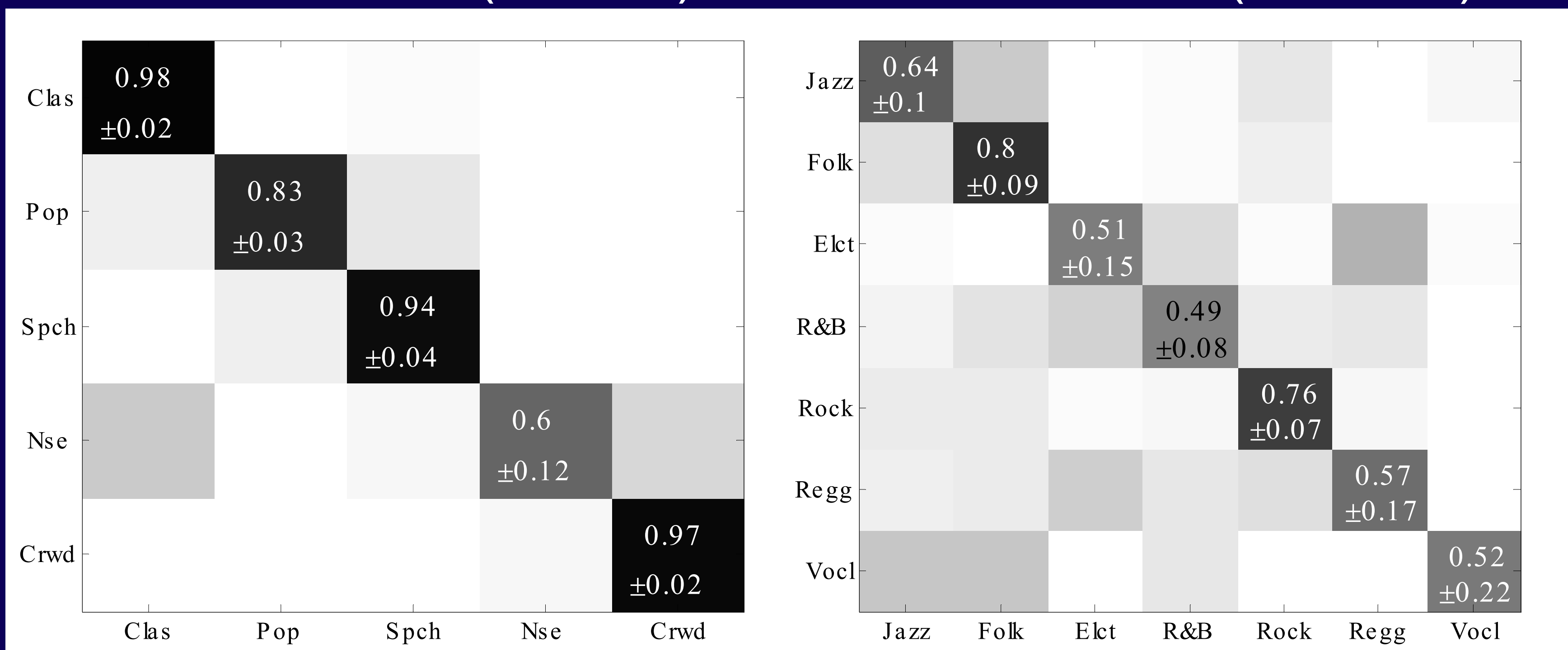
PHILIPS

Results: Standard Low Level features

Classification with 9 best features

General Audio (86±4%) | Music Genre (61±11%)

Real Class



Classification Result

Results: MFCC features

Feature ranking: General Audio, Music Genre

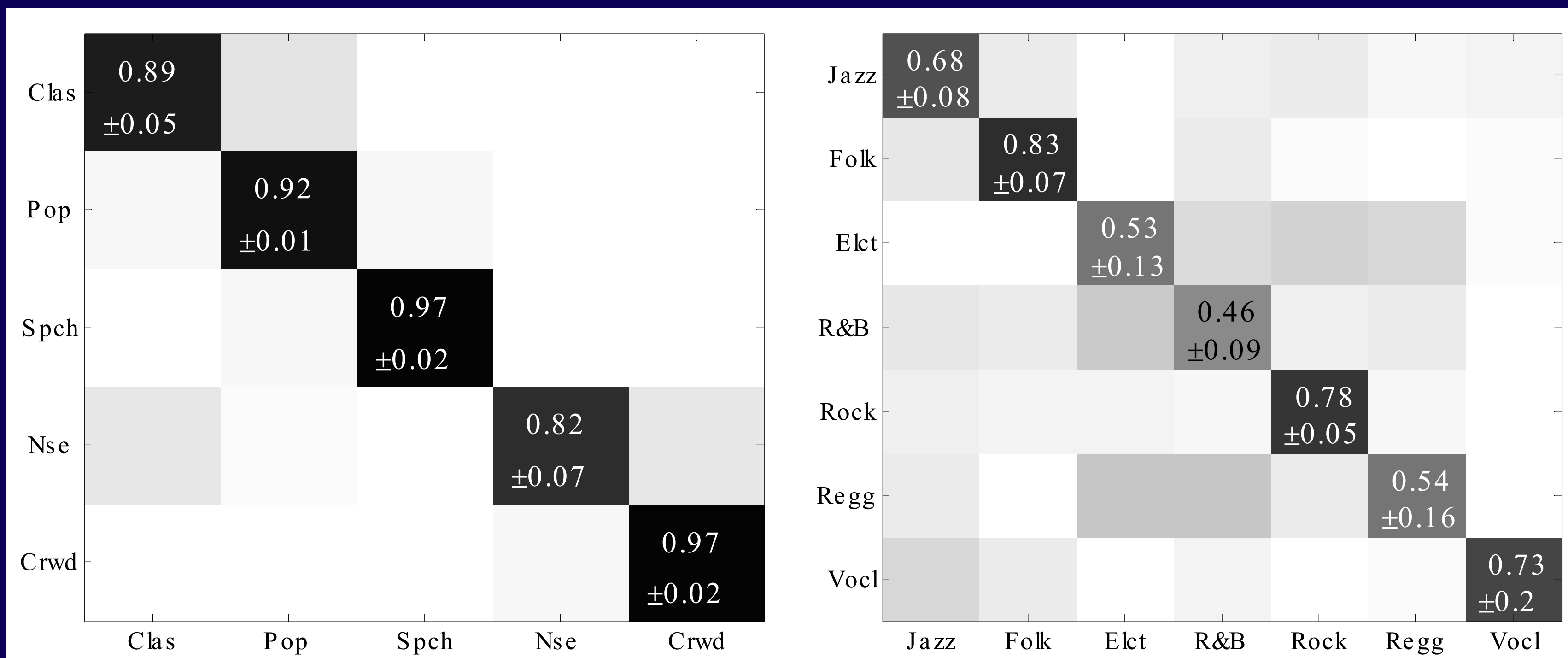
	DC	1-2 Hz	3-15 Hz	20-43 Hz
1. MFCC 0	3, 2		2, 6	1
2. MFCC 1	1, 4			
3. MFCC 2	5, 7			
4. MFCC 3	3			
5. MFCC 4	6			
6. MFCC 5	5			
7. MFCC 6	9			
8. MFCC 7				
9. MFCC 8	7			
10. MFCC 9				4
11. MFCC 10	8, 8			
12. MFCC 11				
13. MFCC 12	9			

Results: MFCC features

Classification with 9 best features

General Audio (92±3%) | Music Genre (65±10%)

Real Class



Classification Result

Results: Psychoacoustic features

Feature ranking: General Audio, Music Genre

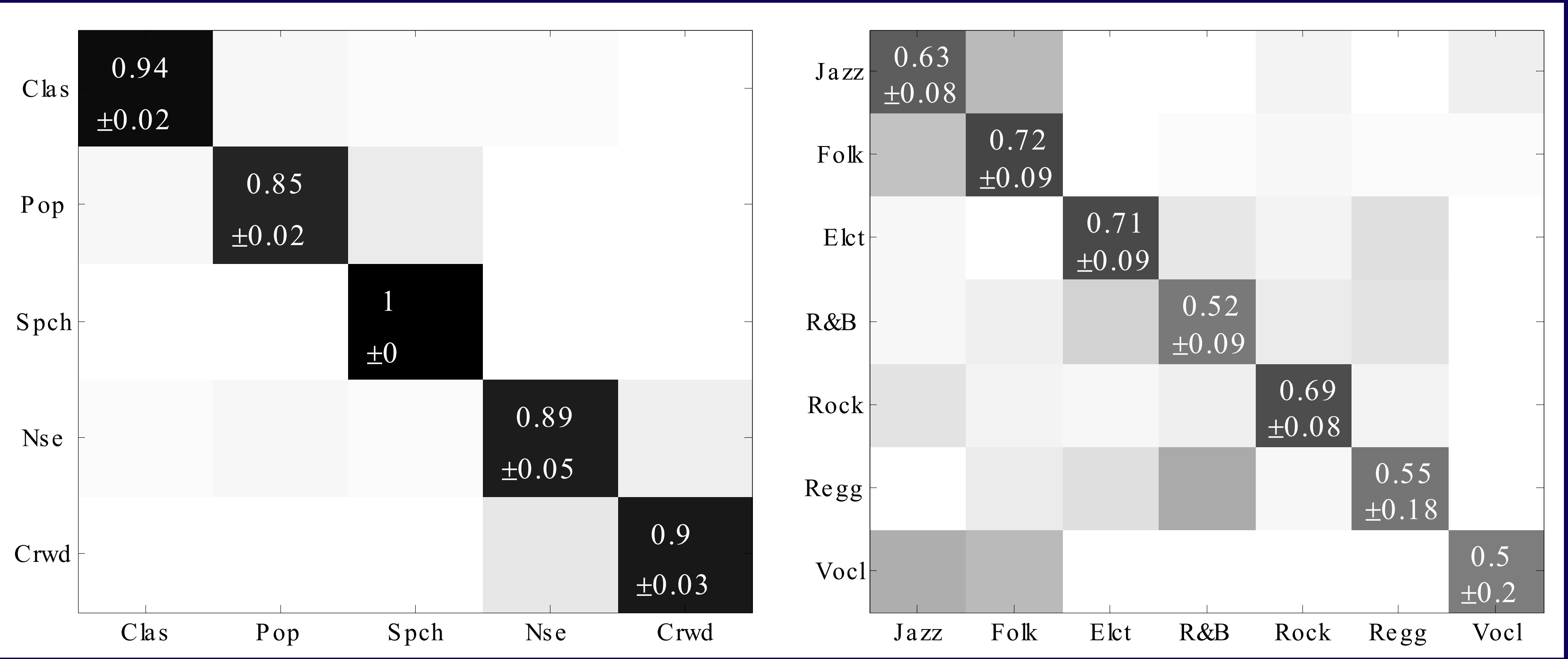
	DC	1-2 Hz	3-15 Hz	20-43 Hz
1. Roughness	3, 2	N/A	N/A	N/A
2. Roughness Std. Dev.	7	N/A	N/A	N/A
3. Loudness	4, 5	8	6, 6	5, 4
4. Sharpness	2, 1	9, 7	1, 3	8, 9

Results: Psychoacoustic features

Classification with 9 best features

General Audio (92±3%)　　Music Genre (62±10%)

Real Class



Classification Result

Results: AFTE features

Feature ranking: General Audio, Music Genre

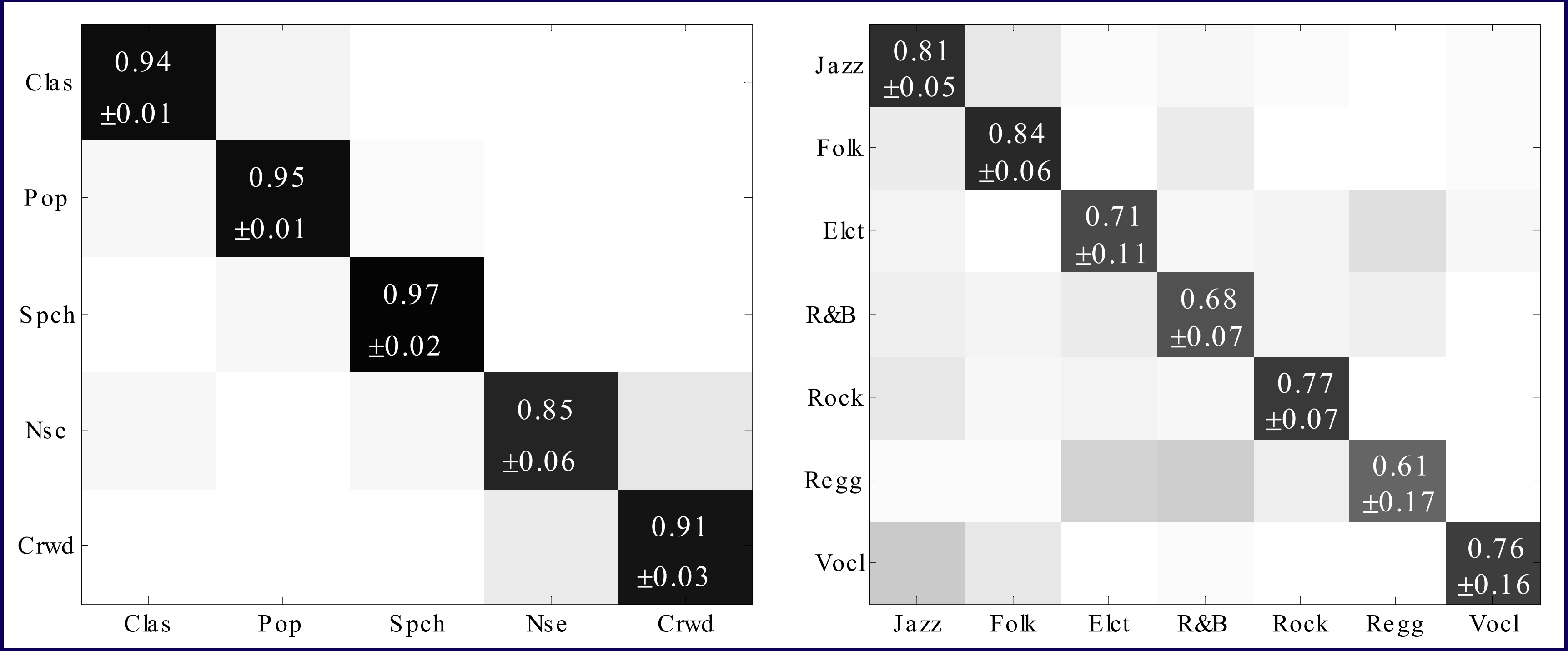
	DC	3-15 Hz	20-150 Hz	150-1000 Hz
1. AFTE 1 (Fc = 26 Hz)	7, 6		N/A	N/A
2. AFTE 2 (Fc = 88 Hz)	1	7	N/A	N/A
3. AFTE 3 (Fc = 164 Hz)	1, 3			N/A
4. AFTE 4 (Fc = 258 Hz)	8			N/A
7. AFTE 7 (Fc = 703 Hz)		5	6	N/A
8. AFTE 8 (Fc = 927 Hz)				N/A
9. AFTE 9 (Fc = 1206 Hz)	4		9	
12. AFTE 12 (Fc = 2514 Hz)	8		9	
16. AFTE 16 (Fc = 6279 Hz)	5			
17. AFTE 17 (Fc = 7848 Hz)				
18. AFTE 18 (Fc = 9795 Hz)	3, 2		4	2

Results: AFTE features

Classification with 9 best features

General Audio (93±2%)   Music Genre (74±9%)

Real Class



Classification Result

Results Summary

	SLL	MFCC	PA	AFTE
General Audio	86±4%	92±3%	92±3%	93±2%
Music Genre	61±11%	65±10%	62±10%	74±9%

Conclusions

- Classification based on features from an auditory model (AFTE) is better than that from other standard feature sets.
- Temporal modulations of features are important for audio and music classification.
- Feature development can improve audio and music classification.

PHILIPS