Music Signal Processing and Applications in Recognition
Athanasia Zlatintsi and Petros Maragos
School of ECE, National Technical University of Athens 15773, Greece
http://cvspl.cs.ntua.gr

1. Outline/Contributions
Analysis of music signals for the extraction of effective descriptors for automatic classification.
- Processing/analysis of music signals with nonlinear methods
  - Fractal theory
  - AM-FM model
  - Introduction of new descriptors based on Bag-of-Words models.
- Experimental evaluation in applications such as:
  - Recognition of musical instruments
  - Recognition of different genres of music
- Study of the AM-FM model for salient event detection and audio summary creation.
- Extension of the proposed model on multiresolution analysis.
- Development of a systematic salience movie database.

2. Motivation
- The power and the role of music in human life from ancient times (Pythagoras, Plato, Aristotle) until today.
- Music’s use in entertainment, advertising, cinema, therapy, teaching, work, cultural heritage, etc.
- The amount of digital music and multimedia data in general.

3. Multiscale Fractal Dimension (MFD):
Quantifies the multiscale complexity of the waveforms, i.e., the degree of its fractalization.

\[ D = \frac{\log(f(\epsilon))}{\log(\epsilon)} \]

Algorithm: based on multiscale nonlinear operators of morphological filtering that create a geometrical covers around the graph of the signal.

4. Amplitude & Frequency Modulation (AM-FM)
- AM-FM model:
  \[ s(t) = \sum_{i} a_i(t) \cos(\phi_i(t)) \]
  instantaneous amplitude

We model each resonance component of music signals as an amplitude and frequency modulated sinusoid (AM-FM signal), and the whole signal as a sum of such AM-FM components.

5. Musical Instruments
- Timbre (instrument specific quality): Quality of sound that distinguishes two sounds of the same pitch, loudness and duration, thus associated with the identification of environmental sound sources.

Example of Average MDFs on Attack of 7 Instruments

- Higher D for small scales
- More fragmented
- Increased value of \( D = 1 \)
- Clear distinction of D among some of the instruments.

6. Conclusions:
1) Classification with MFD
   - Error Rate Reduction (ERR) up to 32%.
2) Classification with AM-FM
   - ERR up to 38% (AM-FM only).
   - ERR up to 60% for 7 instruments (AM-FM fused with MFCC).
3) Iterative-ESA: Possible estimation of the harmonic content of a tone.

7. Audio Summarization
- System Overview: bottom-up processing
- Modulation Features: Energy, Feature Extraction, Audio Salience
- Audio Events: Summary Evaluation: Fusion and for Global Normalization
- Summary Evaluation: for various Fusions and for Global Normalization
- Audio Summary: Freq. and Mean Instantaneous Frequency (FIMF)

8. Publications
- Journal Articles:

- Conference Papers:

- Books and Monographs:

Acknowledgments
This research has been financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) – Research Funding Program: Heracleitus II, Investing in knowledge society through the European Social Fund.