EUSIPCO 2012

AM-FM Modulation Features for Music Instrument Signal Analysis and Recognition

Athanasia Zlatintsi and Petros Maragos
School of ECE, National Technical University of Athens 15773, Greece
http://cvsp.cs.ntua.gr

1. Outline - Contributions

- ☐ **Motivation**: Small fluctuations or micro-modulations occur naturally in both human voice and musical instruments.
- ➡ Exploration of micro-modulations, based on nonlinear **AM-FM modeling** using the **multiband Gabor Energy Separation Algorithm** (Gabor-ESA) [7].
- ☐ AM-FM modulations can capture the fine-structure and the rapid fluctuations of musical signals.
- ☐ Application of the **Iterative-ESA** [5] for the estimation of the center frequencies f_c of the Gabor filterbank.
- ☐ **Goal**: Alternative feature configurations that offer a modest improvement to the performance of a recognition task.

2. Amplitude and Frequency Modulation (AMFM)

- ☐ Musical signals represented as a combination of different "resonances", approximately corresponding to oscillation systems formed by the instruments' characteristics and the sound production procedure.
- ☐ 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.

AM-FM model

$$s(t) = \sum_{i=1}^{K} \underset{\substack{\uparrow \\ \text{instantaneous amplitude}}}{\alpha_i(t)} \cos(\underset{\substack{\uparrow \\ \text{phase}}}{\varphi_i(t)})$$

Energy Separation Algorithm (ESA)

estimates the instantaneous amplitude and frequency

$$|\alpha(t)| \approx \frac{\Psi[x(t)]}{\sqrt{\Psi[\dot{x}(t)]}} \qquad f(t) \approx \frac{1}{2\pi}\sqrt{\frac{\Psi[\dot{x}(t)]}{\Psi[x(t)]}}$$

Teager Energy Operator $\Psi[x] = \dot{x}^2 - x\ddot{x} \qquad (\dot{x} = dx/dt)$

Gabor-ESA

combination of the continuous time ESA and Gabor filtering of the signal (smoother instantaneous estimates)

$$\Psi[\underset{\substack{\downarrow \\ \text{input signal}}}{s(t)} * \underset{\substack{\downarrow \\ \text{Gabor impulse response}}}{g(t)}] = \left[s(t) * \frac{dg(t)}{dt}\right]^2 - (s(t)*g(t))\left[s(t) * \frac{d^2g(t)}{dt^2}\right]$$

Modulation Features

Mean Instantaneous Amplitude:

short-time mean of the $|\alpha_i(t)|$ for each resonance i .

- *Parameterizes the resonance amplitudes and captures part of the nonlinear behavior of the signal.*

Mean Instantaneous Frequency:

short-time weighted mean of the $f_i(t)$.

- *Models the time-varying frequency of the resonance.*
- *Provides information about the signal's fine structure taking advantage of ESA's excellent time resolution.*

3. Experimental Configuration and Results

Experimental configuration

- Input signal: Fs = 44.1 KHz.
- Dynamic range: piano to forte.
- **12 Gabor Filters** with mel placing and 50% overlap. Optimum joint time-frequency resolution.
- Analysis on 30 ms segments, updated every 15 ms.

Experimental Setup

- ☐ 2070 notes, 12 instruments
- ☐ 5 feature sets
  1. 24 AMFM (12 m-IAM + 12 m-IFM)
  2. 72 AMFM_Δ
  3. AMFM_{50} (18 m-IAM + 32 m-IFM) after PCA
  4. AMFM_{39} (12 m-IAM + 27 m-IFM) after PCA
  5. 39 MFCC_Δ

  - $\text{AMFM}_\Delta + \text{MFCC}_\Delta$
  - $\text{AMFM}_{39} + \text{MFCC}_\Delta$
  - $\text{AMFM}_{50} + \text{MFCC}_\Delta$

- ☐ 2 classification methods
  - GMM M[1-3] mixtures
  - Multi-stream HMM N[3-9] states, M[1-3] mixtures

Results

MFCC-AMFM

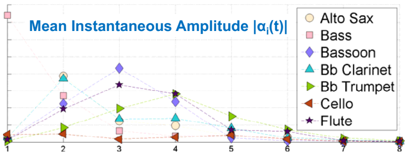
- ☐ Error Rate Reduction **(ERR) ≈ 56% (AMFM_{39})**
- ☐ Enhanced discrimination except for: Bass, Bass Trombone & Oboe
- ☐ Improves: Bass Trombone, Horn, Clarinet

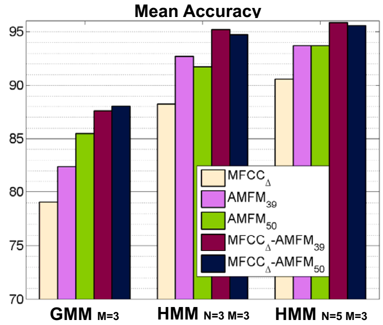
AM-FM

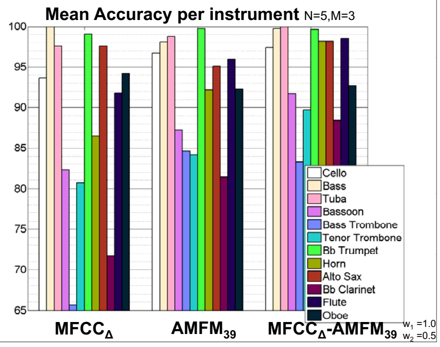
- ☐ **ERR ≈ 15%, 38%, 33% (AMFM_{39})**
- ☐ Better recognition on all, except Bass, Saxophone & Oboe
- ☐ High recognition: Trumpet, Tuba, Bass & Cello
- ☐ Low recognition: Clarinet & Flute, Bass & Tenor Trombone

MFCC

- ☐ High recognition: Bass, Trumpet, Tuba & Sax
- ☐ Low recognition: Clarinet, Flute & Oboe, Bass & Tenor Trombone, Bassoon & Horn



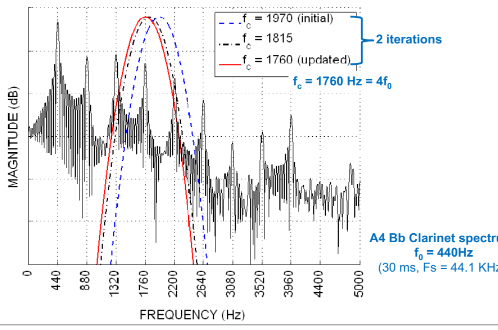




4. Iterative-ESA

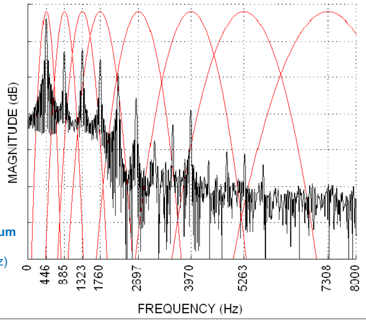
- ☐ Iterative application of ESA to the Gabor filtered signal.
- ☐ Reduces the importance of having good initial estimates of the filterbank's center frequencies, since they are iteratively adjusted after every iteration.
- ☐ f_c: tend to converge on frequencies close to integer multiples of the fundamental frequency, i.e., the harmonics.
- ☐ Some Gabor filters favored to converge at the same center frequency.
- ➢ Possible estimation of the harmonic content of a tone.

Estimation of the 5th Gabor filter



A4 Bb Clarinet spectrum
f_0 = 440Hz
(30 ms, Fs = 44.1 KHz)

Estimated center frequencies f_c



References

[1] P. Maragos, J. F. Kaiser and T. F. Quatieri, "Energy Separation in signal modulations with application to speech analysis", *IEEE Trans. Signal Process.*, vol. 41, pp. 3024–3051, Oct. 1993.
[2] H.M. Hanson, P. Maragos and A. Potamianos, "A system for finding speech formants and modulations via Energy Separation," *IEEE Trans. Speech and Audio Process.*, vol. 2, no. 2, 1994.
[3] A. Potamianos and P. Maragos, "Speech formant frequency and band- width tracking using multiband energy demodulation", *J. Acoust. Soc. Amer.,* vol. 9, no. 6, pp. 196–200, June 1996.
[4] D. Dimitriadis, P. Maragos, and A. Potamianos, "Robust AM-FM features for speech recognition", *IEEE Signal Processing Letters*, vol. 12, no. 9, pp. 621–624, Sep. 2005.
[5] A. Potamianos and P. Maragos, "Speech processing applications using an AM-FM Modulation model", *Speech Communication*, vol.28, pp.195–209, 1999.
[6] J. McAulay and T.F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 34, no. 4, 1986.
[7] X. Serra, "Musical sound modeling with sinusoids plus noise", *Musical Signal Processing*, Pope S. Picialli A. Roads, C. and G. (Eds.) De Poli, Eds. Swets & Zeitlinger, 1997.
[8] A.S. Bregman, Auditory Scene analysis, *The perceptual organization of sound*, MIT Press: Cambridge, MA, 1990
[9] University of Iowa Musical Instrument Sample Database. [Online]. Available: http://theremin.music.uiowa.edu/index.html

Acknowledgments

This research has been co-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.

More information: Please contact: *[nzlat,maragos]@cs.ntua.gr*

European Union
European Social Fund
OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION, LIFELONG LEARNING AND RELIGIOUS AFFAIRS
MANAGING AUTHORITY
Co- financed by Greece and the European Union
NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND