AN AMPLITUDE- AND FREQUENCY-MODULATION VOCODER FOR AUDIO SIGNAL PROCESSING

Sascha Disch, Bernd Edler

Laboratorium für Informationstechnologie
Leibniz Universität Hannover

Overview

- Motivation
- Fields of application
- Modulation analysis/synthesis
- Modulation processing
- Spectrograms/audio samples
  - Modulation detail modification
  - Music transposition
  - Key mode conversion
- Summary

lfi Laboratorium für Informationstechnologie
Leibniz Universität Hannover

Motivation

*From a **perceptual** view, audio signals are composed of low bandwidth and low frequency sub-processes, which **modulate** much higher carrier frequencies.*

- Modulation decomposition?
- Goal: intuitive access to perceptually relevant parameters (pitch, roughness, timbre)
- Ill-defined: Infinite number of possible decompositions
- Border conditions of a suitable decomposition:
  - Carriers as ‚sketch' representation => **signal adaptive**
  - Modification at minimal artifacts => **perceptually meaningful**
  - Fine grain scalability towards ‚*transparency*' => **seamless** transition from **parametric** to **waveform** representation
  - Appropriate synthesis

lfi Laboratorium für Informationstechnologie
Leibniz Universität Hannover

Fields of application

- Audio effects (studio, music production)
  - Pitch manipulation
  - Tremolo, Vibrato (slow variable: < 20Hz)
  - Roughness (fast variable: > 20 Hz) [14]-[16]
- Efficient audio coding methods
  - Fine grain scalable audio coding [1]
  - Bandwidth replication [20]
- Semantic audio signal processing
  - Source separation using modulation features (‚common fate')
- Hearing aids [9]
- Research tool
  - Evaluate influence of fine vs. coarse structure in human auditory perception via creation of „audio chimaeras" [11]

lfi Laboratorium für Informationstechnologie
Leibniz Universität Hannover

Modulation analysis/synthesis: principle

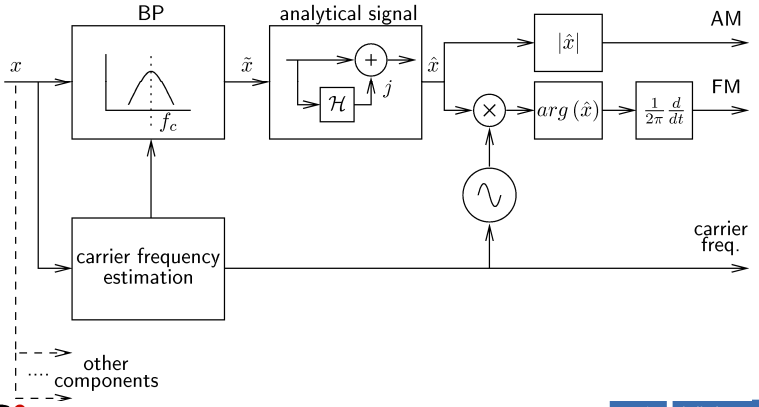
- Time-blockwise decomposition into sets of sub-band coefficients, each comprising
  - Carrier
  - Amplitude modulation (AM)
  - Frequency modulation (FM)
- Sub-bands centered at carriers in local spectral **centers of gravity (COG)**

=> Spectral segmentation task

- **Signal adaptive** alignment of sub-bands
- **Perceptual** criteria (critical bands)
- **Seamless coverage** of spectrum by sum of all sub-bands

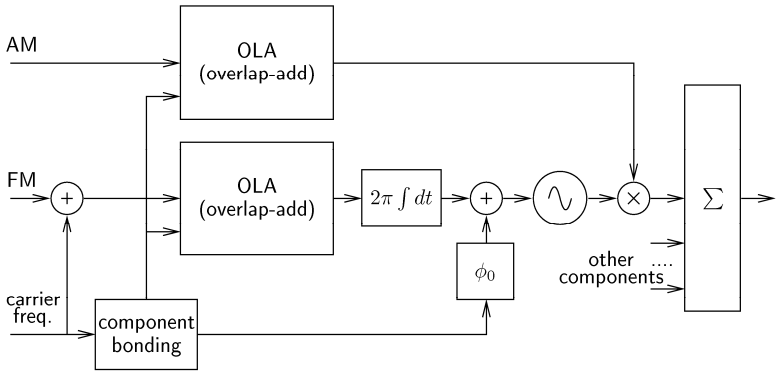
Modulation analysis/synthesis: system (1/2)

- Analysis: analytical sub-band signals
  - AM: absolute value (envelope)
  - FM: instantaneous frequency (IF) from derivative of phase



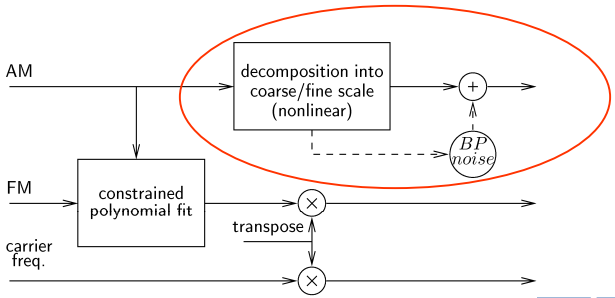
Modulation analysis/synthesis: system (2/2)

- Synthesis
  - Component bonding for carrier phase continuity
  - Overlap-add in modulation component domain

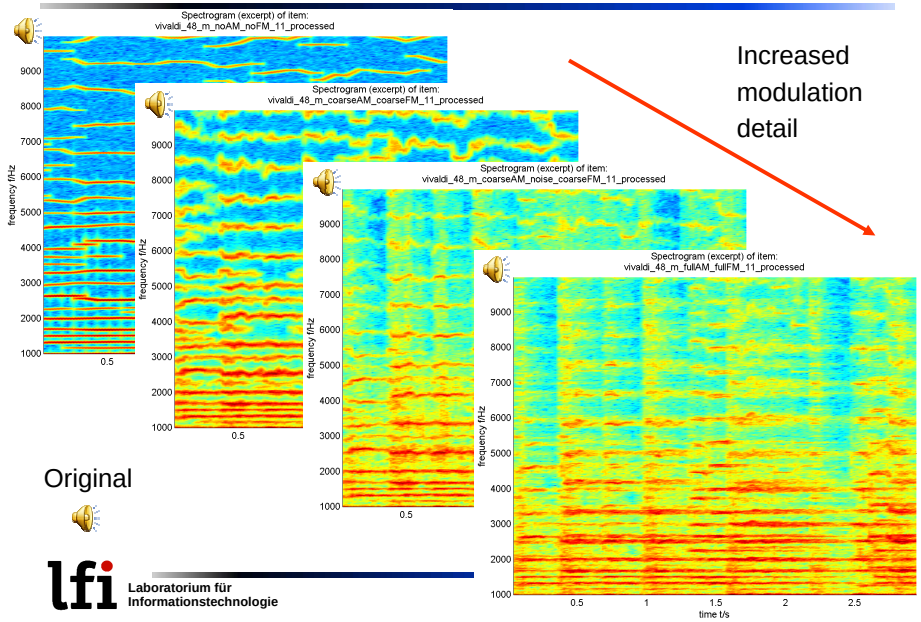


Modulation processing

- Separation of coarse and fine amplitude modulation
  - Coarse AM (=> ADSR envelope parameter "***A**ttack, **D**ecay, **S**ustain, **R**elease*"), "*tremolo*"
  - Fine AM (noise, auditory roughness)
  - Substitution of fine AM with band-limited noise weighted by coarse envelope (generalization of **PNS** [**p**erceptual **n**oise **s**ubstitution])

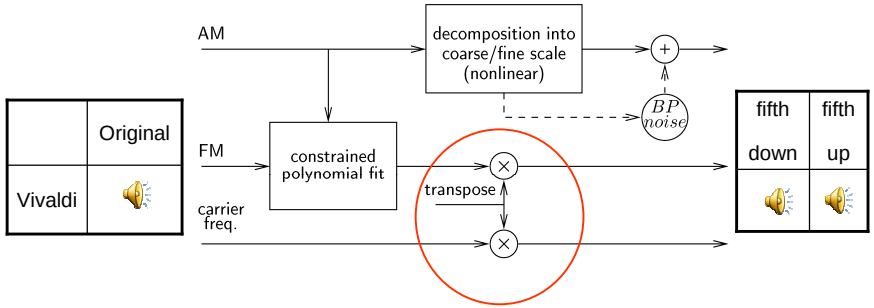


Results: modulation detail modification



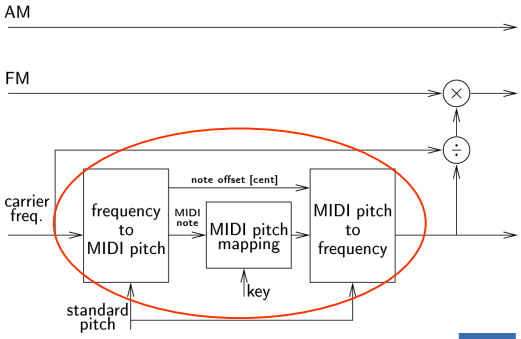
Results: music transposition

- Changed pitch at unchanged playback speed
- Stretch/squeeze of global spectrum
  - Scaling of all carrier frequencies and FM by constant factor
  - Temporal structure of signal is maintained since it is contained in the (unchanged) AM

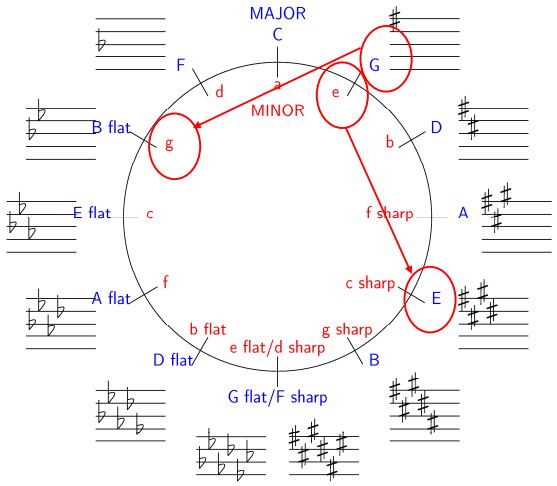


Results: key mode conversion (1/3)

- Key conversion of polyphonic music, e.g. major <-> minor
- Spectral mapping of selected carrier frequencies
  - Mapping of carriers onto MIDI-pitch representation
  - A-priori information: standard pitch and key
  - Shift of selected pitch values according to scale change



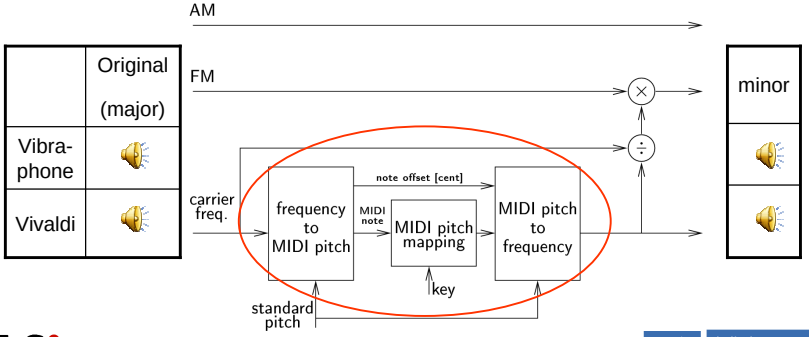
Results: key mode conversion (2/3)



- Mapping of tones according to scale change
- Circle of fifth
- Rules:
  - **Major to minor**: 3 steps counter clockwise
  - **Minor to major**: 3 steps clockwise
- 1 Step == mapping of one tone

Results: key mode conversion (3/3)

- Key conversion of polyphonic music, e.g. major <-> minor
- Spectral mapping of selected carrier frequencies
  - Mapping of carriers onto MIDI-pitch representation
  - A-priori information: standard pitch and key
  - Shift of selected pitch values according to scale change



Summary

- A promising novel method for audio decomposition into sub-band modulation components
- Possible link between waveform and parametric coding
- Spectral decomposition of paramount importance for perceptual quality
- High quality synthesis
- Scalable modulation detail, straightforward handle on pitch
- Applications scenarios
  - Audio codec/bandwidth extension
  - Research tool for auditory perception
  - New exciting audio effects

lfi Laboratorium für Informationstechnologie
Leibniz Universität Hannover

End

questions/discussion

lfi Laboratorium für Informationstechnologie
Leibniz Universität Hannover

References

(1) M. Vinton and L. Atlas, "A Scalable And Progressive Audio Codec," in *Proc. of ICASSP 2001*, pp. 3277-3280, 2001
(2) H. Dudley, "The vocoder," in *Bell Labs Record*, vol. 17, pp. 122-126, 1939
(3) J. L. Flanagan and R. M. Golden, "Phase Vocoder," in *Bell System Technical Journal*, vol. 45, pp. 1493-1509, 1966
(4) J. L. Flanagan, "Parametric coding of speech spectra," *J. Acoust. Soc. Am.*, vol. 68 (2), pp. 412-419, 1980
(5) U. Zoelzer, *DAFX: Digital Audio Effects*, Wiley & Sons, pp. 201-298, 2002
(6) H. Kawahara, "Speech representation and transformation using adaptive interpolation of weighted spectrum: vocoder revisited," in *Proc. of ICASSP 1997*, vol. 2, pp. 1303-1306, 1997
(7) A. Rao and R. Kumaresan, "On decomposing speech into modulated components," in *IEEE Trans. on Speech and Audio Processing*, vol. 8, pp. 240-254, 2000
(8) M. Christensen et al., "Multiband amplitude modulated sinusoidal audio modelling," in *IEEE Proc. of ICASSP 2004*, vol. 4, pp. 169-172, 2004
(9) K. Nie and F. Zeng, "A perception-based processing strategy for cochlear implants and speech coding," in *Proc. of the 26th IEEE-EMBS*, vol. 6, pp. 4205-4208, 2004
(10) J. Thiemann and P. Kabal, "Reconstructing Audio Signals from Modified Non-Coherent Hilbert Envelopes," in *Proc. Interspeech (Antwerp, Belgium)*, pp. 534-537, 2007
(11) Z. M. Smith and B. Delgutte and A. J. Oxenham, "Chimaeric sounds reveal dichotomies in auditory perception," in *Nature*, vol. 416, pp. 87-90, 2002
(12) J. N. Anantharaman and A.K. Krishnamurthy, L.L Feth, "Intensity weighted average of instantaneous frequency as a model for frequency discrimination," in *J. Acoust. Soc. Am.*, vol. 94 (2), pp. 723-729, 1993
(13) O. Ghitza, "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," in *J. Acoust. Soc. Amer.*, vol. 110(3), pp. 1628-1640, 2001
(14) E. Zwicker and H. Fastl, *Psychoacoustics - Facts and Models*, Springer, 1999
(15) E. Terhardt, "On the perception of periodic sound fluctuations (roughness)," in *Acustica*, vol. 30, pp. 201-213, 1974
(16) P. Daniel and R. Weber, "Psychoacoustical Roughness: Implementation of an Optimized Model," in *Acustica*, vol. 83, pp. 113-123, 1997
(17) P. Loughlin and B. Tacer, "Comments on the interpretation of instantaneous frequency," in *IEEE Signal Processing Lett.*, vol. 4, pp. 123-125, 1997.
(18) D. Wei and A. Bovik, "On the instantaneous frequencies of multicomponent AM-FM signals," in *IEEE Signal Processing Lett.*, vol. 5, pp. 84-86, 1998.
(19) Q. Li and L. Atlas, "Over-modulated AM-FM decomposition," in *Proceedings of the SPIE*, vol. 5559, pp. 172-183, 2004
(20) M. Dietz, L. Liljeryd, K. Kjörling and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," in *112th AES Convention*, Munich, May 2002.
(21) Qin Li and Les Atlas, "Coherent Modulation Filtering For Speech," in Proc. of ICASSP 2008, pp. 4481-4484, 2008

lfi Laboratorium für Informationstechnologie
Leibniz Universität Hannover