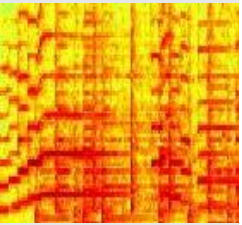


Music Recognition

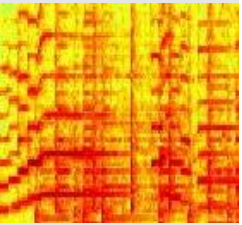
Extracting A Monophonic Instrument Out Of A Jazz Quartett Recording

Term Thesis

Alain Brenzikofer

June 5, 2004

Aim Of The Project



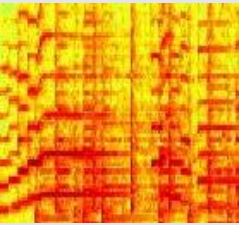
Given

- stereo recording of a jazz quartett (ts, p, b, dr)
- a-priori knowledge about instruments and their spectral/temporal characteristics

extract following information:

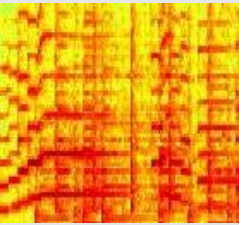
- is a saxophone playing?
- if yes $\Rightarrow$ prepare information for transscription

Signal Analysis

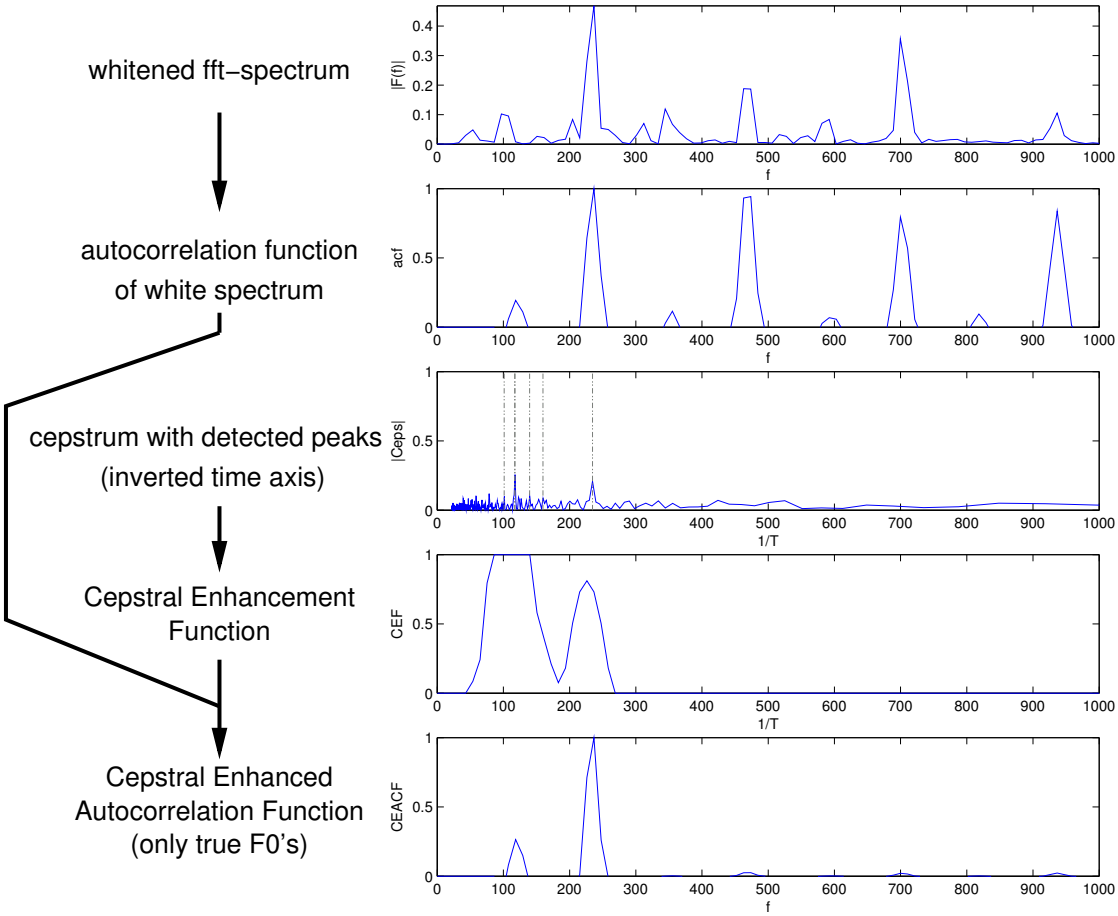


To extract multiple F0-trajectories, a special method has been developed.
Main problems:

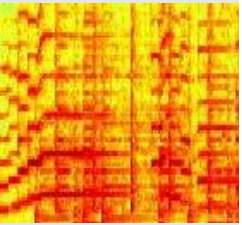
- overlapping of harmonics destroys information
- instruments playing in octave interval are hard to distinguish
- time-frequency uncertainty is critical when using temporal features as well



Multiple-F0-Detection

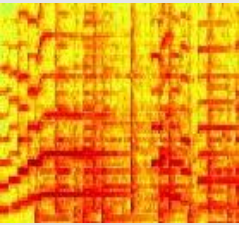


Time Domain Agents

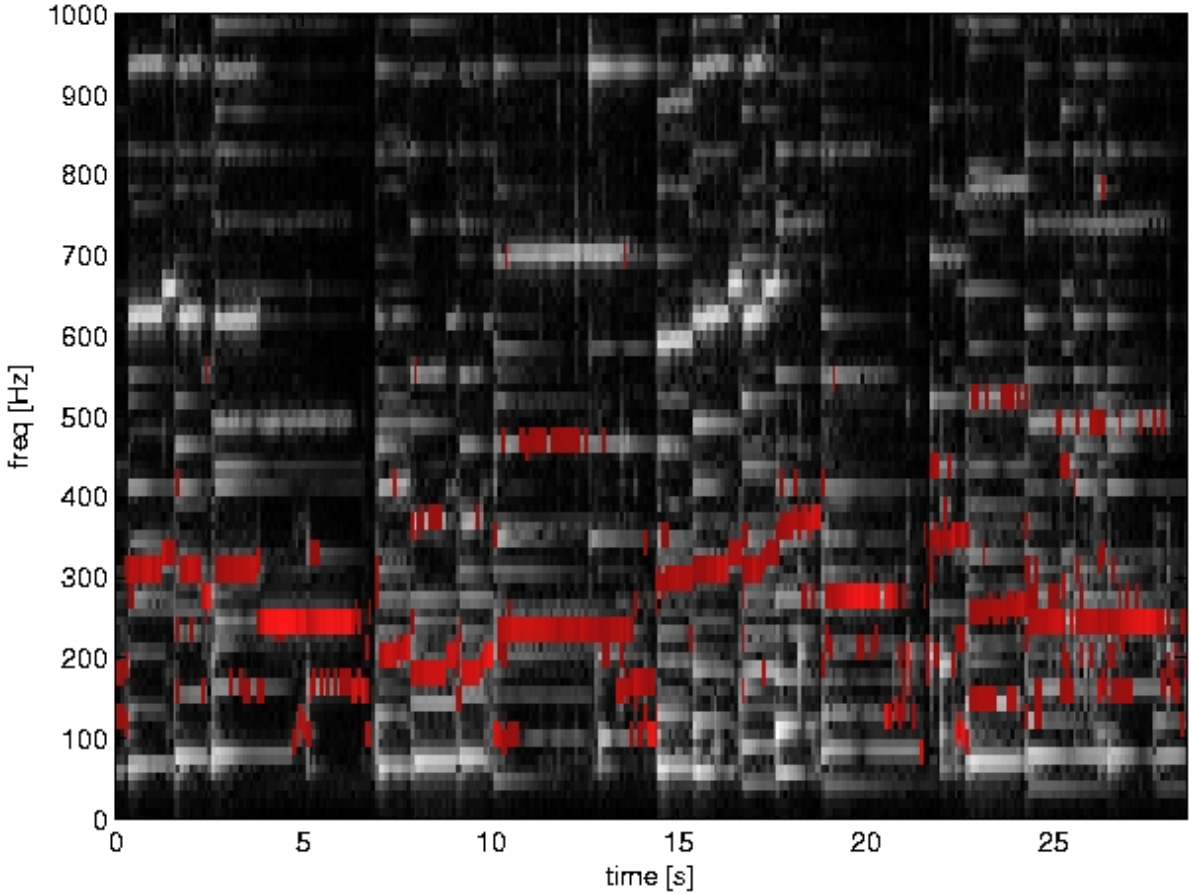


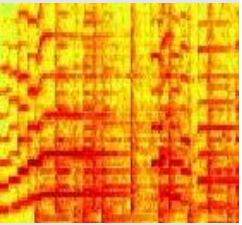
Agents group detected F0's to notes

- agents follow the notes until they end or change by at least a semitone.
- a new agent is generated whenever a new possible F0 trajectory starts (for one instrument)
- more than one agents per detected F0 is possible
- only agents with a minimal duration and signal energy are used
- agents can be extended in time to catch true beginning and ending



Example F0 Recognition Output





Training

Data

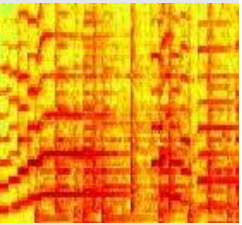
	seconds of music	spectral feat.vect.	temporal feat.vect.
sax:	106	1032	117
nosax:	142	1493	346

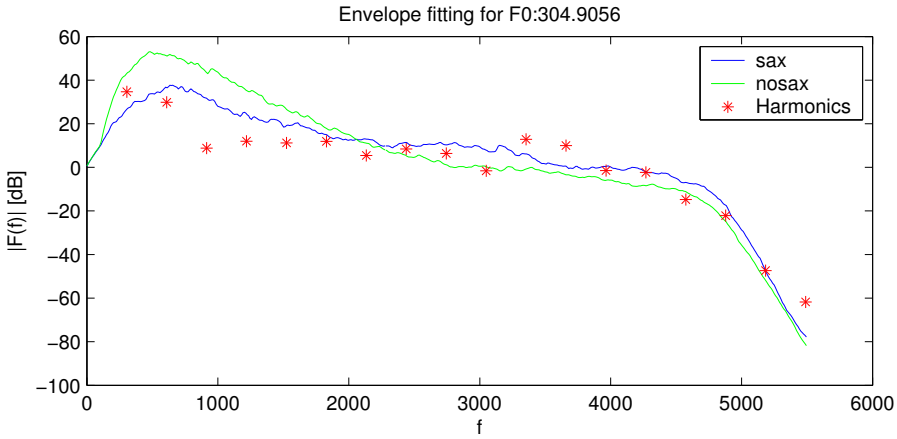
Features

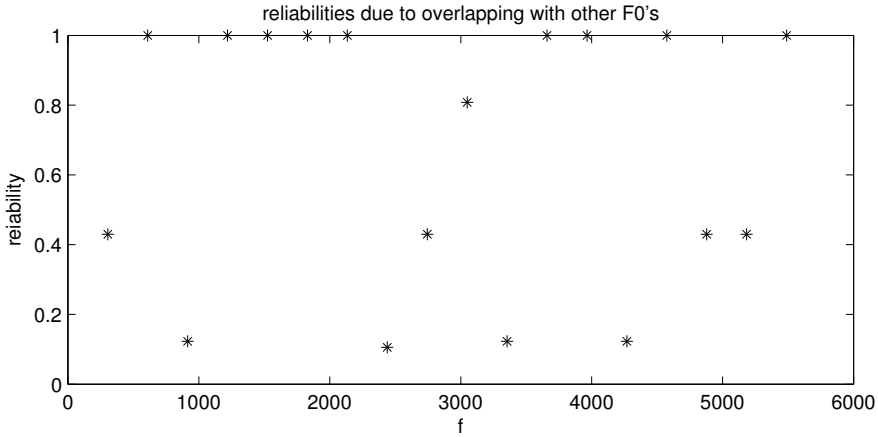
Two separate groups of feature vectors: spectral and temporal. Spectral features can be calculated for every window, temporal features only for every recognized note.

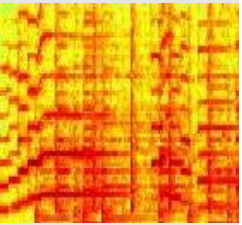
Spectral Features (1) Harmonics Spectral Envelope

Amplitudes of all harmonics are compared to a model envelope



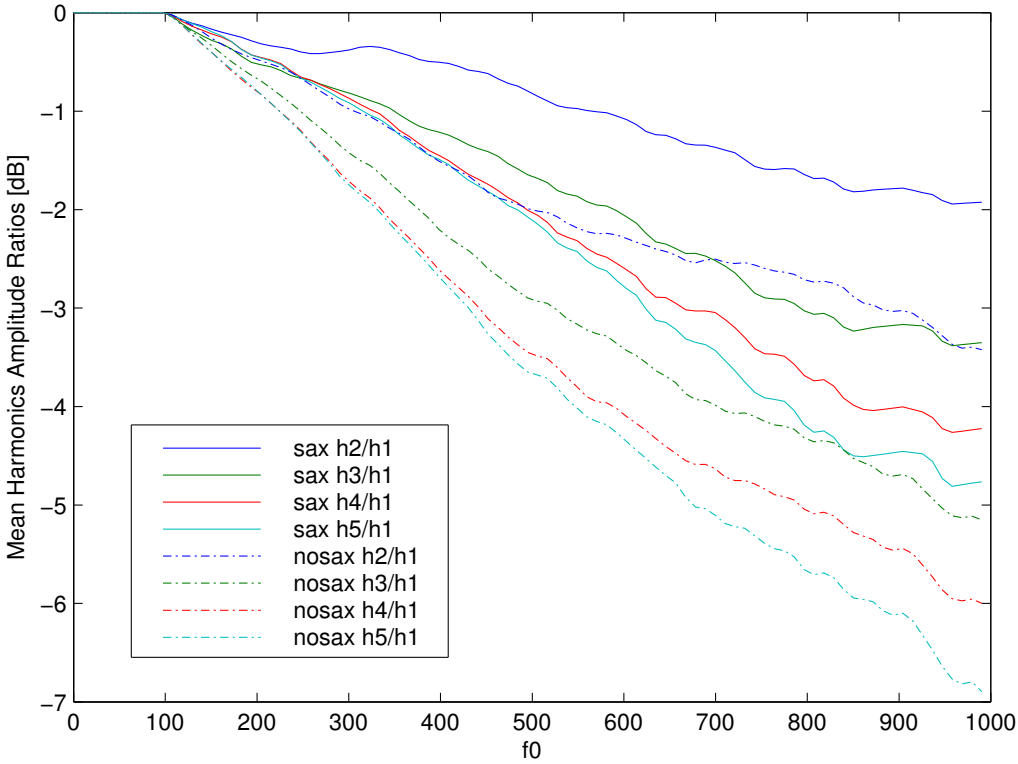




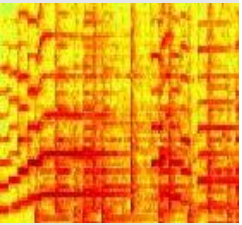


Spectral Features (2) Harmonic Ratios Envelope

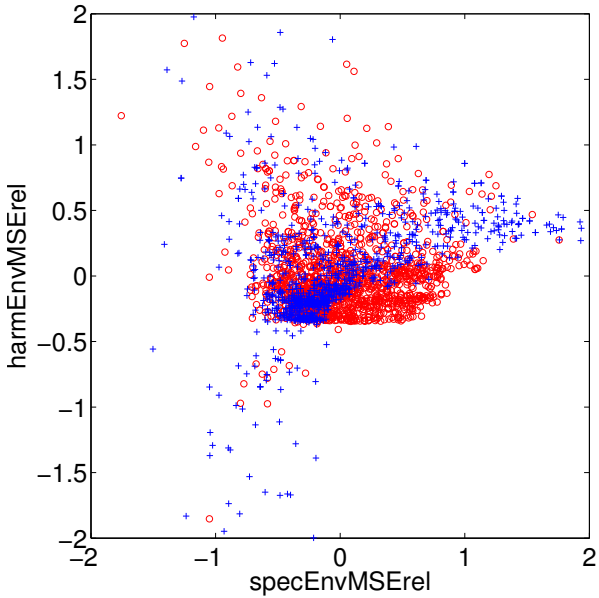
independent of F0 frequency, $\frac{\text{harmonics 2-5 amplitudes}}{\text{F0 amplitude}}$ is calculated and compared to a model envelope, depending on F0 (very noisy feature)



Spectral Features - Data

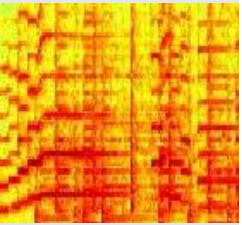


MSE of harmonics spectral envelope and MSE of harmonics proportions envelope



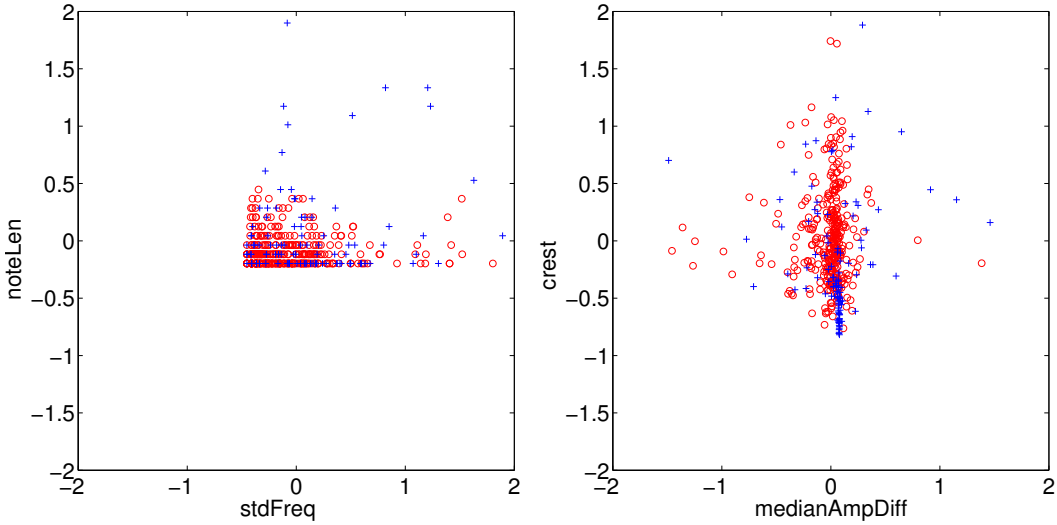
sax:blue, nosax: red

Temporal Features



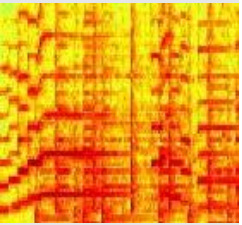
typical saxophone envelope:

typical piano/bass envelope:



sax:blue, nosax: red

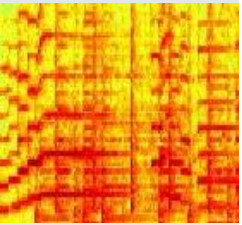
Training Algorithms



- **GMM:** Two models for each class (spectral features: 50 gaussians, temporal features: 10 gaussians). Spectral and temporal likelihoods have equal weight for classification.

  $\Rightarrow$ performed best

- **SVM** One spectral, one temporal model. Performed ok but less robust than GMM
- **NN** One spectral, one temporal network. No well-learning net could be found, but performance is similar to above models



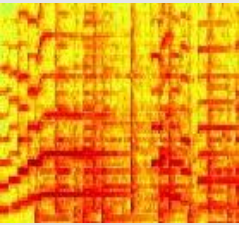
Performance Tests (GMM)

Non-Mixed Music

Evaluation of performance for one-class samples. (can be polyphonic for nosax-class)

	correctly classified notes	false alarms
saxophone only	34	7
piano only	12	5
bass only	10	0
piano & bass	139	15
overall performance	88%	

Mixtures of all instruments

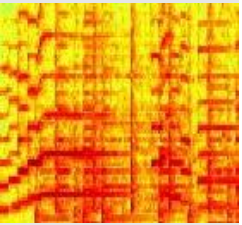


– studio recording with clear, predominant saxophone playing slow melody

	is sax	is nosax	% correct
classified as sax	18	5	78%
classified as nosax	8	11	57%

– low quality live recording with noise

	is sax	is nosax	% correct
classified as sax	7	8	46%
classified as nosax	22	24	52%



Example Classification Output