USING INSTRUMENT RECOGNITION TO EXTRACT MELODIES FROM COMPLEX AUDIO

Jana Eggink and Guy J. Brown
Department of Computer Science
University of Sheffield
{j.eggink, g.brown}@dcs.shef.ac.uk

INTRODUCTION

- Aim: identify the solo instrument in accompanied sonatas and concertos
- Use this knowledge to extract the melody line
- Useful for: automatic music information retrieval, e.g. transcription, 'query-by-humming' systems, automatic indexing and analysis

OVERVIEW

- Instrument recognition and melody extraction without placing any restrictions on the background by focusing on the spectral peaks of the harmonic series of the target tones
- Melody estimation based on multiple F0 candidates
- Observation likelihoods for F0 candidates estimated using Gaussian classifiers trained on the F0s of music with known melody lines
- Different features for observation likelihood of F0 candidates
  - Explicit: absolute frequency + strength + instrument recognition likelihood
  - Implicit: frequency and power of individual partials
- Transition probabilities, i.e. frame based interval likelihoods, estimated from the same training material
- Viterbi search for most likely melody line based on normalised and scaled observation likelihood and transition probabilities

RESULTS

- 5 different solo instruments: flute, clarinet, oboe, violin, cello
- Accompaniment: piano, cembalo or orchestra
- All examples from commercially available classical CDs

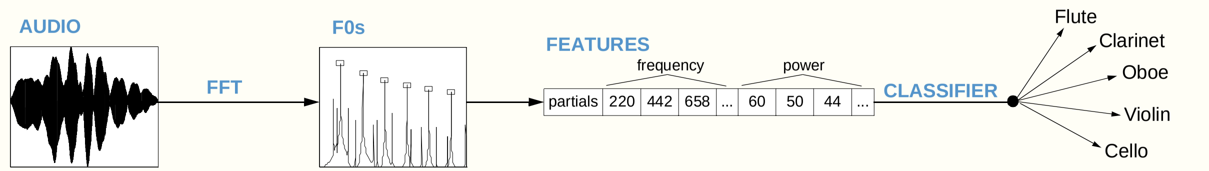
INSTRUMENT RECOGNITION

- 90 different examples, no knowledge about true F0s, accuracy for instrument recognition per piece: 86% (as good as other systems designed to work with monophonic music only)
- 20 examples, 1 minute each manually labeled F0s, 51% of individual frames correct, 95% when pooled and averaged per example

MELODY EXTRACTION

- No pruning, every F0 on a half tone scale between C2 (65 Hz) and C7 (2093 Hz) considered, 61 F0 candidates per frame
- For observation likelihood implicit features based on frequency and power of spectral peaks outperformed explicit coding of frequency+strength+instrument likelihood, despite using only 3 pieces for training
- Observation likelihood and transition probabilities normalised to have the same mean and standard deviation, for best results influence of observation likelihood strengthened (to the power of 4)
- Instrument specific observation likelihoods outperformed those trained on examples from all solo instruments → initial solo instrument recognition is important!
- Most common error: octave doublings of longer sections played in the lower range of the solo instrument
- Number of frames with correct F0, cross validation for 20 examples of 1 minute each: strongest F0: 45% | observation likelihood: 60% | observation likelihood + transitions: 70%

INITIAL INSTRUMENT RECOGNITION



INPUT
- Sound file (windowed)
- Unknown solo instrument + piano or orchestra

F0 ESTIMATION
- Find most prominent harmonic series based on spectral peaks
- Using pattern matching based on harmonic sieves

FEATURES
- Exact frequency position and normalised log-compressed power of every partial of current F0

CLASSIFIER
- Gaussian mixture models (GMMs)
- One GMM for every pitch of every instrument
- Trained on isolated tones with known F0s and solo phrases with F0s estimated by the system

MELODY EXTRACTION

INPUT
- Sound file (windowed)
- Known solo instrument + piano or orchestra

F0 CANDIDATES
- Currently no pruning, every F0 on a musical halftone scale is considered
- 61 candidates per frame

OBSERVATION LIKELIHOODS

VITERBI SEARCH
- Observation likelihood and transition probabilities normalised to same mean and standard deviation
- Influence of observation likelihood strengthened (to the power of 4)

MELODY

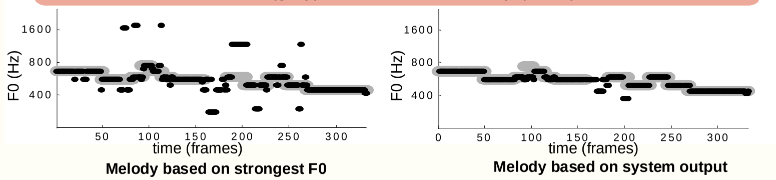
MELODY TONE-MODELS
Explicit features:
- GMMs trained on absolute frequency of F0 + F0 strength + instrument likelihood (as used for initial instrument recognition)
- Numerous training examples for instrument recognition

Implicit features:
- GMMs trained on frequency position and power of every partial of current F0 candidate
- Similar to features used for instrument recognition, but power once normalised to highest value within the harmonic series (as for instrument recognition), and once to highest value within the current frame (to code relative strength of F0 candidate)
- Only 3 training pieces with the same solo instrument

TRANSITION PROBABILITIES
- Interval transitions on a frame-to-frame basis
- Very high probability to stay on same F0, otherwise slightly favouring small intervals of half and full tones, thirds and fifths

TRAINING MATERIAL
- CD recordings of accompanied solo instrument with manually labeled F0s
- Using only pieces with the same solo instrument as the test file

Analysis of the beginning of Mozart's clarinet concerto, showing manually annotated F0s (gray) and estimated melody (black)



ACKNOWLEDGEMENTS
Jana Eggink acknowledges the financial support provided through the European Community's Human Potential Programme under contract HPRN-CT-2002-00276, HOARSE. Guy J. Brown is supported by EPSRC grant GR/R47400/01.