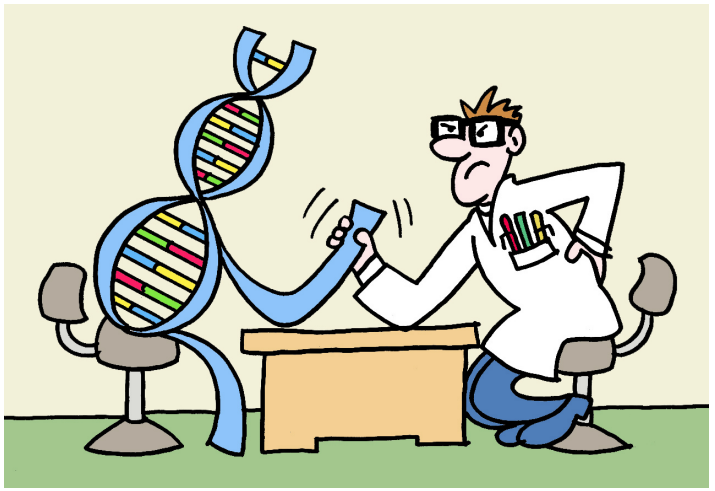
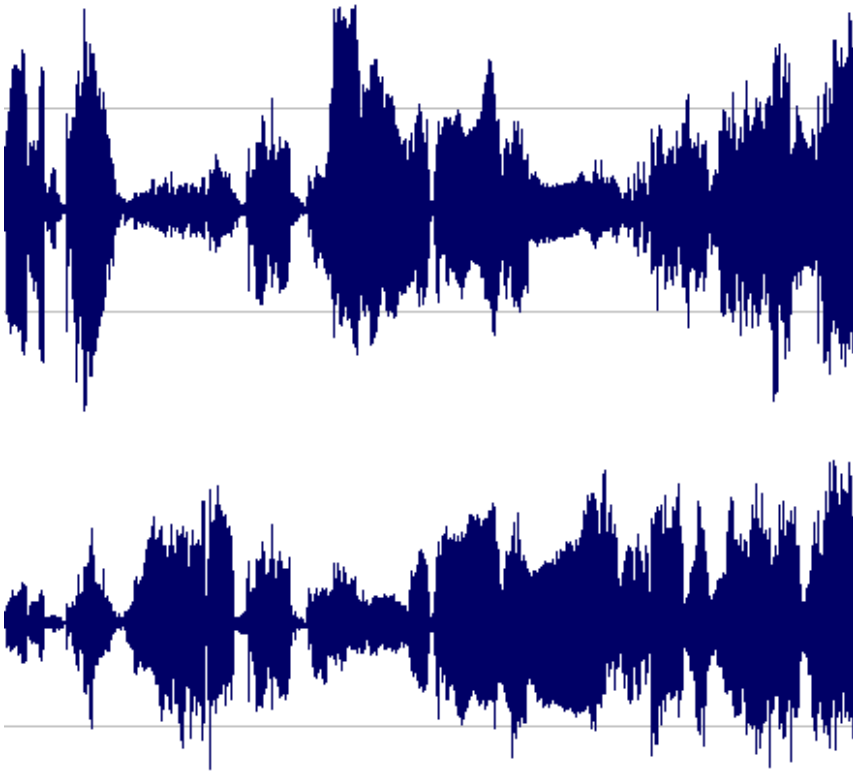


Automatic Transcription of Polyphonic Piano Music using Genetic Algorithms, Adaptive Spectral Envelope Modeling and Dynamic Noise Level Estimation

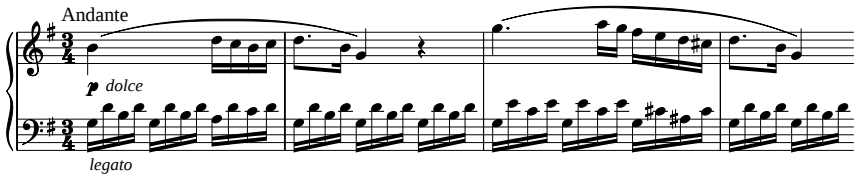
Gustavo Reis
School of Technology and Management, Polytechnic Institute of Leiria, Portugal
Francisco Fernandéz
University of Extremadura, Spain
Aníbal Ferreia
University of Porto, Portugal

Automatic Transcription of Music

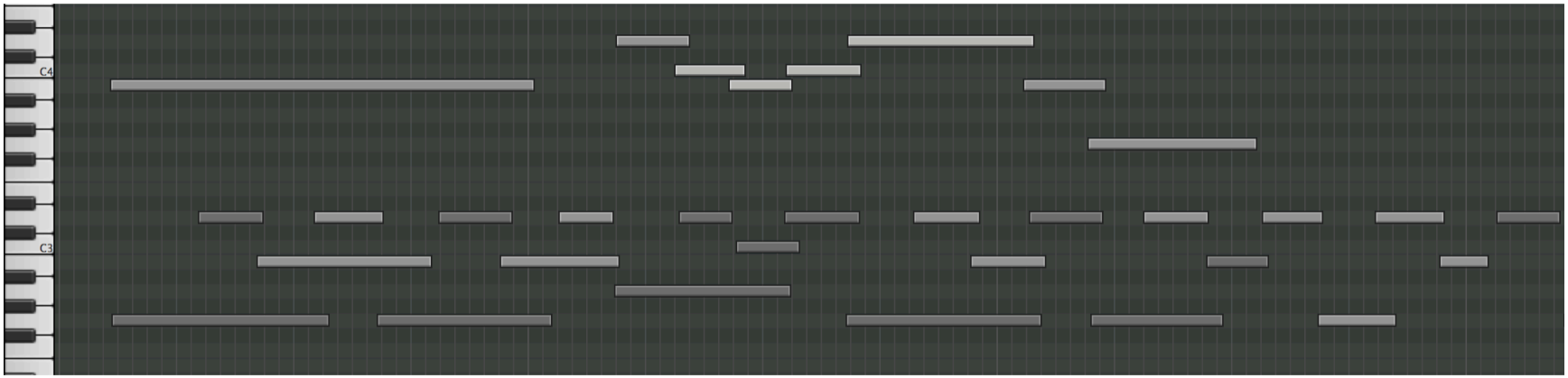
Audio Input



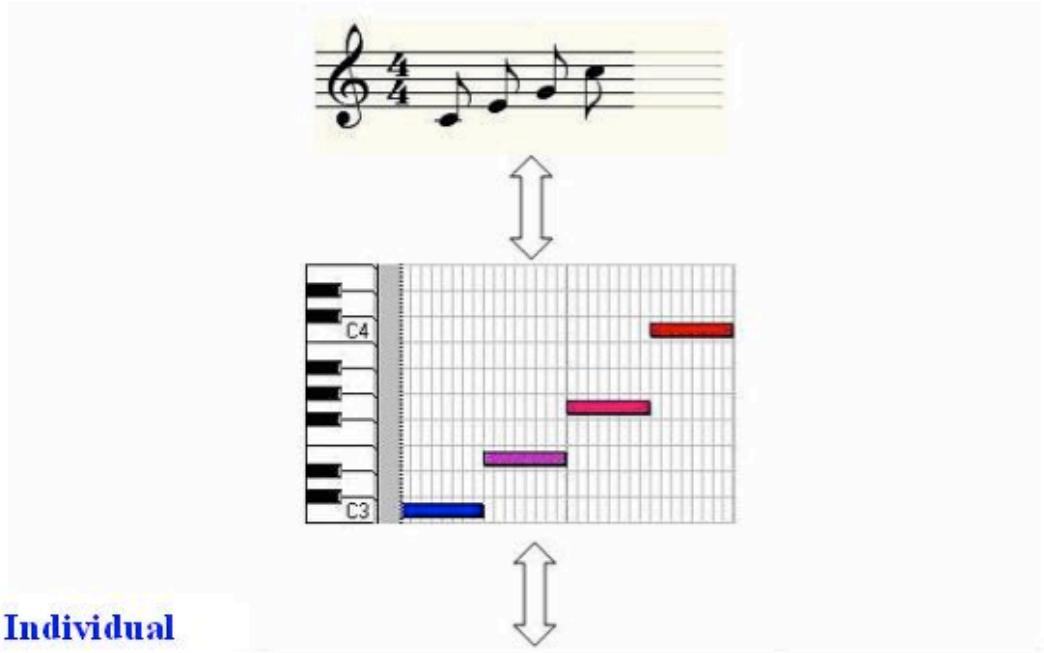
Generated Output



Generated Output: piano-roll



Individual – Candidate transcription



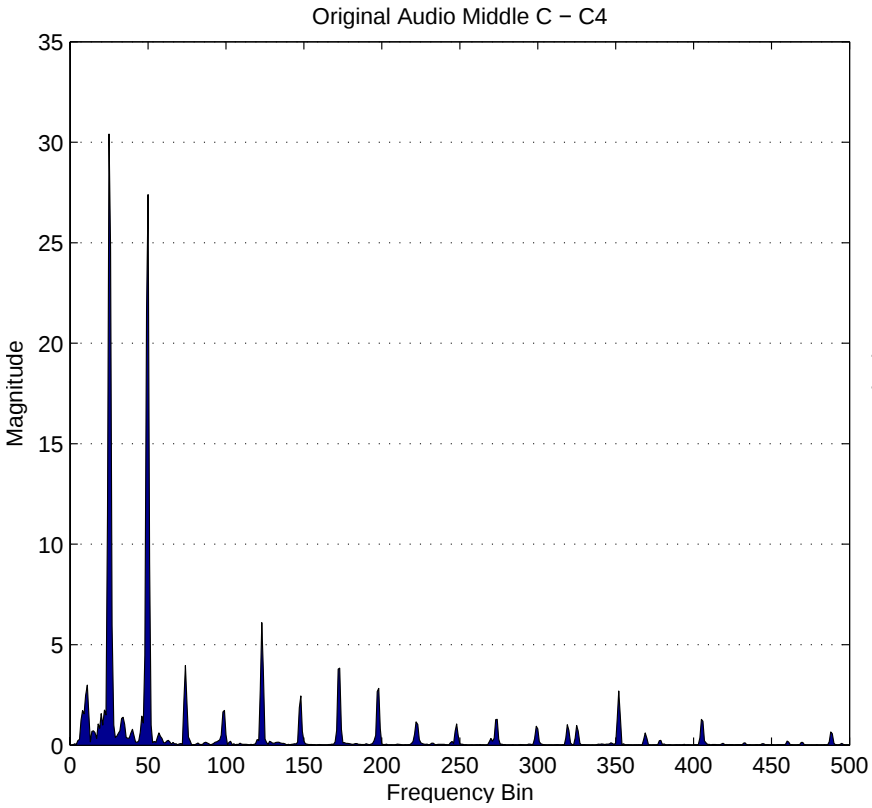
Individual

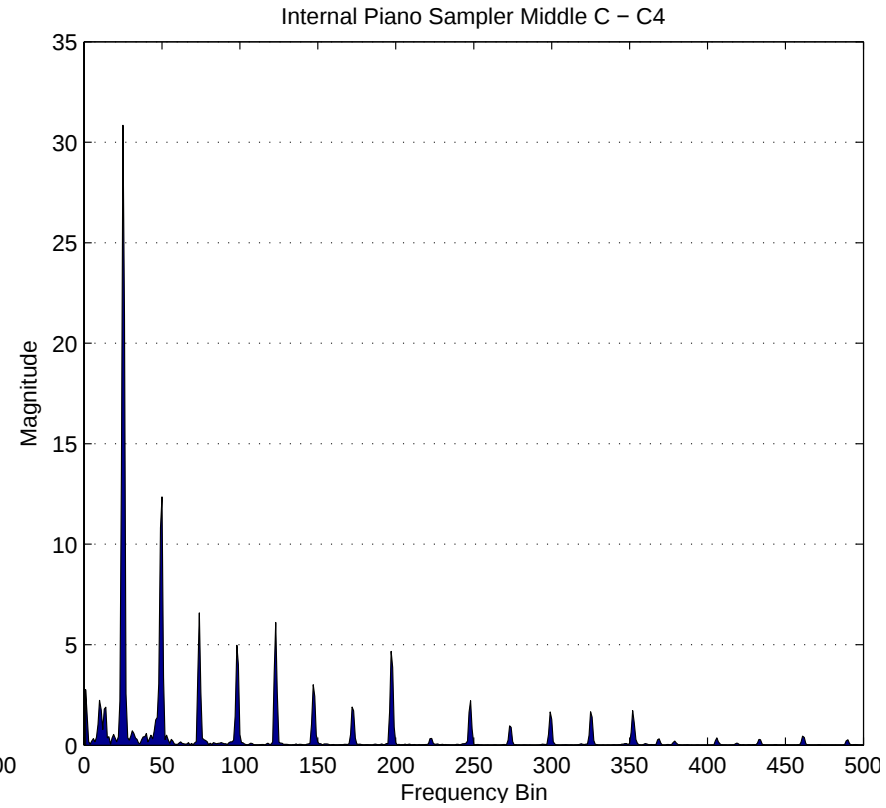
Gene	Gene	Gene	Gene
Note: 60 Start: 0 Duration: 22050 Velocity: 32	Note: 64 Start: 22050 Duration: 22050 Velocity: 64	Note: 67 Start: 44100 Duration: 22050 Velocity: 96	Note: 72 Start: 66150 Duration: 22050 Velocity: 127

Fitness Evaluation

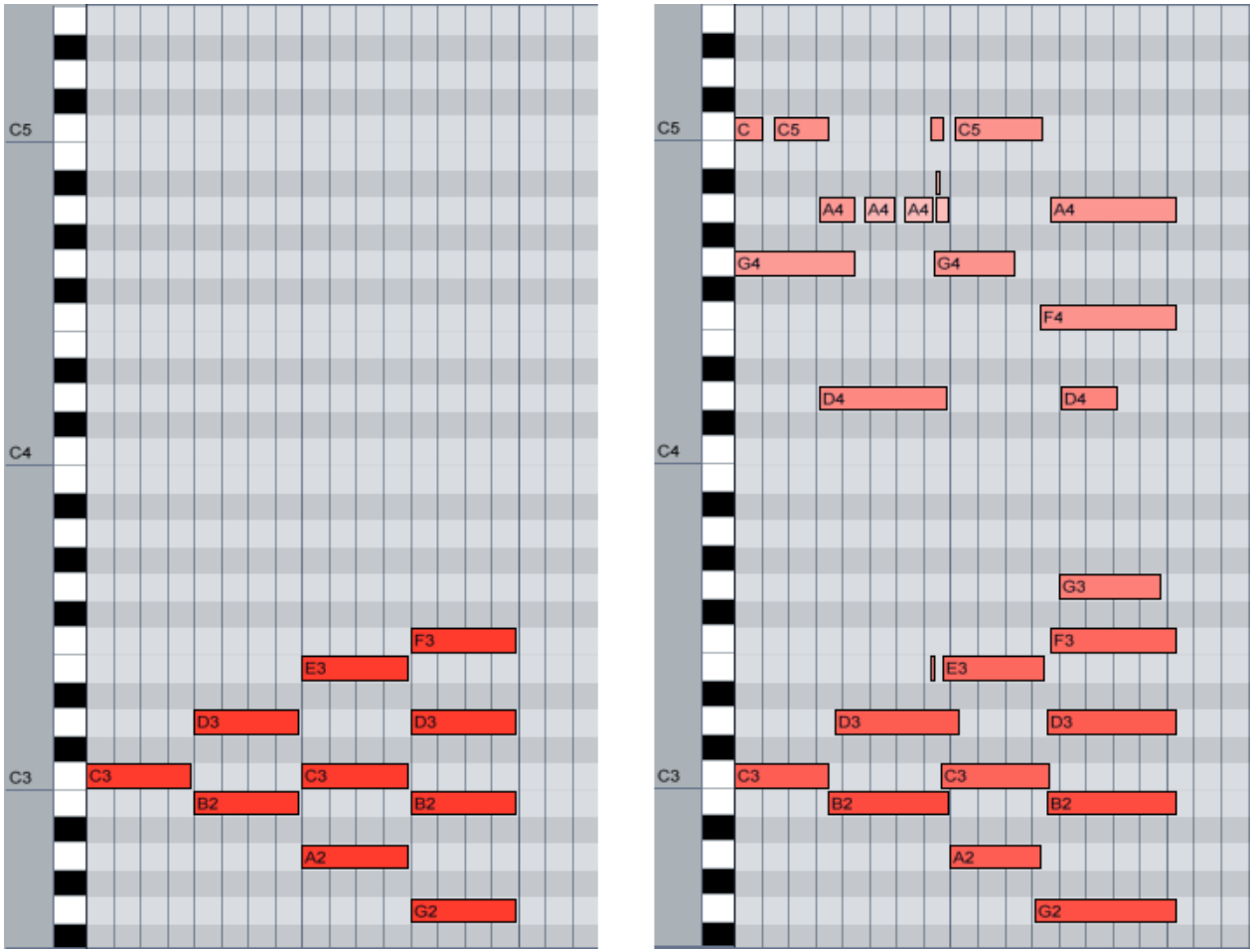
- Each Individual is a score
- Scores do not have sound
- How to perform evaluation?
- Solution: use a synthesizer
  - By using a synthesizer we can convert each candidate transcription into a sound signal and compare to the input signal
- *Sparseness!!!*

Harmonic *Overfitting*

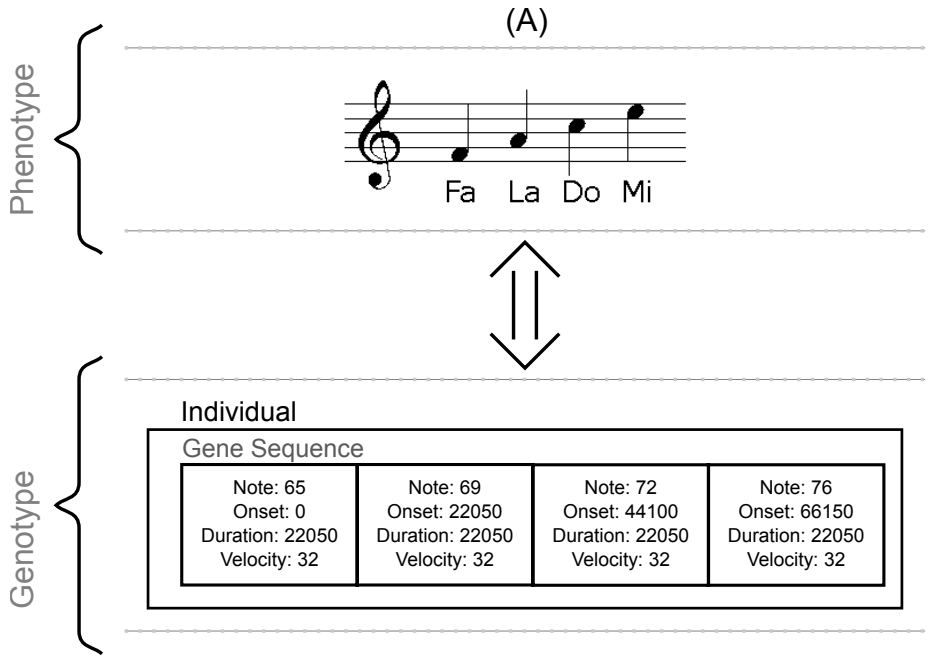


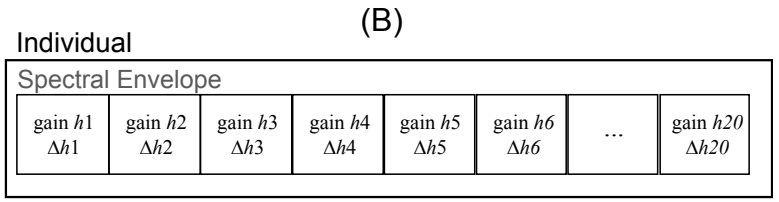


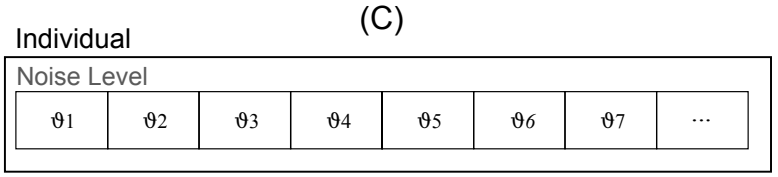
Harmonic *Overfitting*

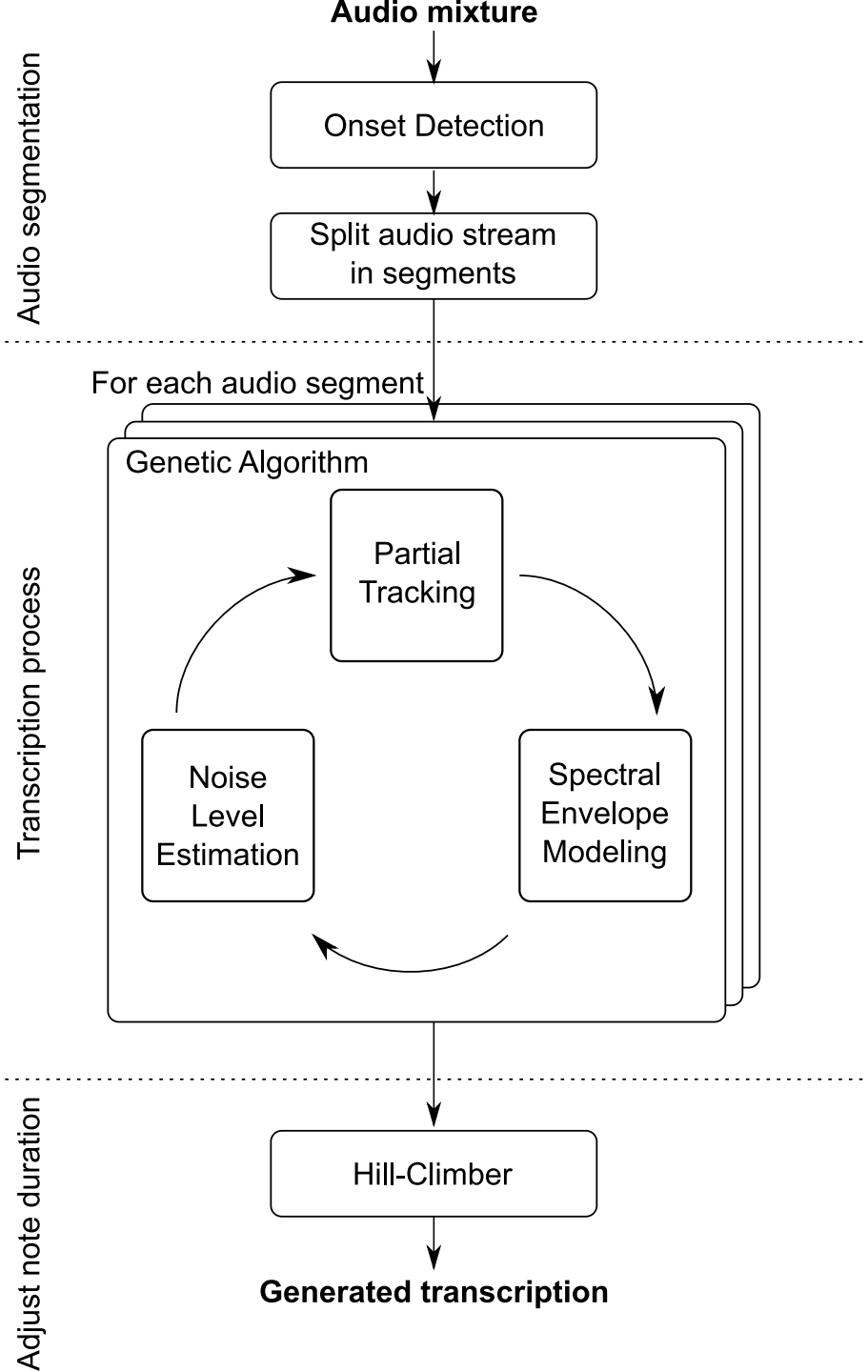


Proposed Solution

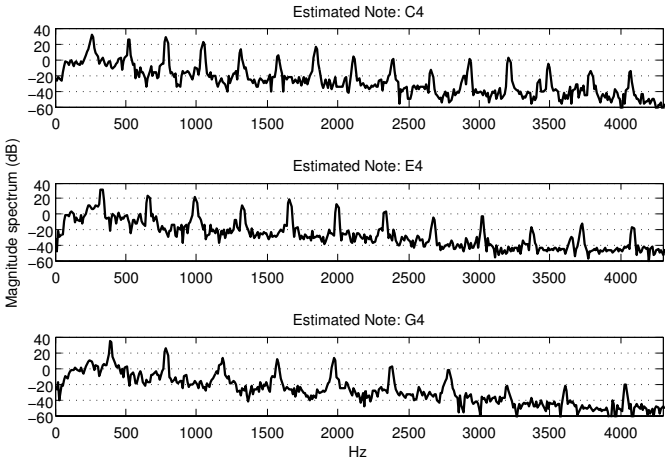


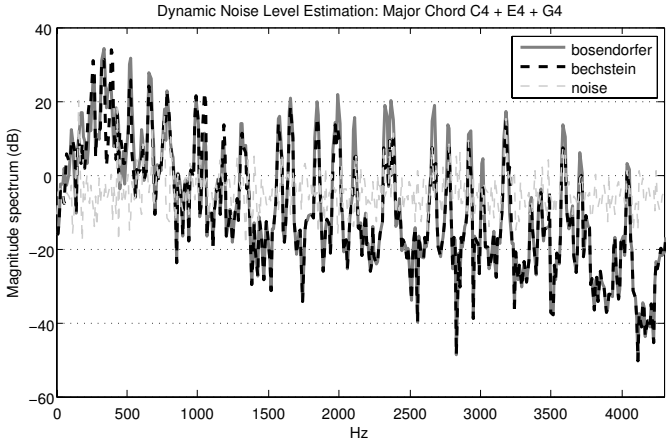


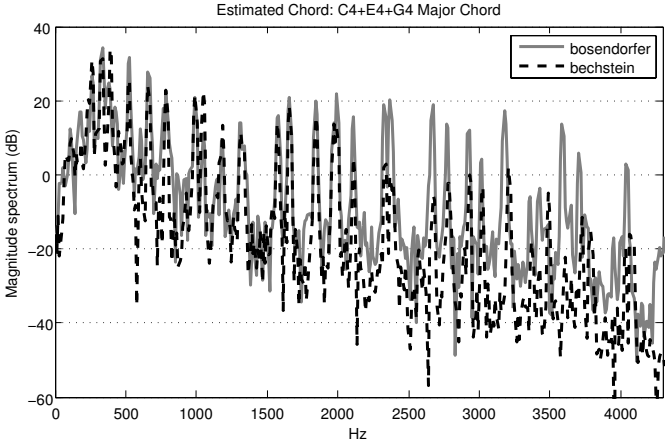


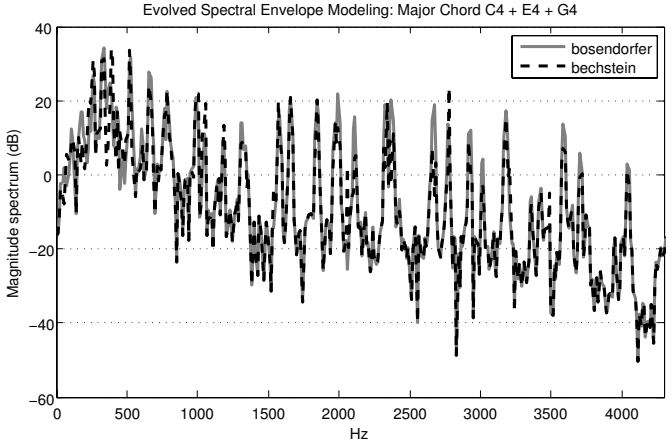

Audio mixture
Audio segmentation
Onset Detection
Split audio stream in segments
For each audio segment
Transcription process
Genetic Algorithm
Partial Tracking
Noise Level Estimation
Spectral Envelope Modeling
Adjust note duration
Hill-Climber
Generated transcription

Transcripition of “C Major” chord









Why is Human Competitive?

- On 3 different benchmarks our approach ranked two times as **2nd best** and one as the **best algorithm** among the state-of-the-art.
- According to the metric that best correlates with the **human hearing perception** our algorithm ranked as the **best**.

Why is Human Competitive? Results – Mirex 2011

Ave. F-Measure Onset-Offset Piano Subtask	
YR1	0.2127
RFF1	**0.1941**
LYC1	0.1926
YR3	0.1913
RFF2	0.1550
BD3	0.1136
BD2	0.1003

Why is Human Competitive?

- More results in: http://www.estg.ipleiria.pt/~gustavo.reis/benchmark/

Why our Approach is the Best?

- Our algorithm also mimics how musicians learn to play a tune by hear:
  - the algorithm first listens the audio,
  - then tries to play it the best as it can,
  - improving the generated tune from iteration to iteration.
- According to the human hearing perception:
  - the algorithm tries to reproduce the sound that has eared.
- This way, our algorithm, **besides behaving like humans**, it **also outperforms their algorithms**, according to the human hearing perception.

Thank you a lot!