Ph.D. Thesis Proposal

Automatic Construction of Synthetic Musical Instruments and Performers

Ning Hu

Carnegie Mellon University

December 2004

Thesis Committee

- **Roger B. Dannenberg, *Chair***
- **Michael S. Lewicki**
- **Richard M. Stern**
- **David Wessel (UC Berkeley)**

Roadmap

- **Introduction**
- **System Structure**
- **Main Modules**
  - **Audio Alignment and Segmentation**
  - **Instrument Model**
  - **Performance Model**
- **Schedule**
- **Conclusion**

Introduction

- **Define thesis topic**
- **Thesis statement**
- **Start with modeling *the trumpet* for *classical music***
- **Contributions of the thesis**
- **Criteria for success**

Topic Definition

- ***Title***

  Automatic Construction of Synthetic Musical Instruments and Performers

- ***Meaning***
  - **Framework that builds music synthesis**

Topic Definition

- ***Title***

  <u>Automatic Construction</u> of Synthetic Musical Instruments and Performers

- ***Meaning***
  - **Framework that builds music synthesis**
  - **Automatic construction**

Topic Definition

- ***Title***

  Automatic Construction of Synthetic Musical Instruments and Performers

- ***Meaning***
  - **Framework that builds music synthesis**
  - **Automatic construction**
  - **Instrument model**

Topic Definition

- ***Title***

  Automatic Construction of Synthetic Musical Instruments and Performers

- ***Meaning***
  - **Framework that builds music synthesis**
  - **Automatic construction**
  - **Instrument model**
  - **Performance model**

Topic Definition

- ***Title***

  Automatic Construction of Synthetic Musical Instruments and Performers

- ***Meaning***
  - **Framework that builds music synthesis**
    - **High quality**
    - **High performance**
    - **Capable of modeling different instruments**
    - **Capable of modeling different music styles**

Topic Definition

- ***Title***

  Automatic Construction of Synthetic Musical Instruments and Performers

- ***Meaning***
  - **Automatic construction**
    - **Use machine learning techniques**
    - **Learn from performance examples**
    - **Constructs instrument model and performance model**

Topic Definition

- ***Title***

  Automatic Construction of Synthetic Musical Instruments and Performers

- ***Meaning***
  - **Instrument model**
    - **Similar to traditional concept of *synthesis***
    - **Input: control signals**
    - **Output: synthesized sound samples**

Topic Definition

- ***Title***

  Automatic Construction of Synthetic Musical Instruments and Performers

- ***Meaning***
  - **Performance model**
    - **Generating appropriate control signals from music context is crucial**
    - **Drives instrument model**
    - **Input: digital score (music notation)**
    - **Output: control signals**

Topic Definition

- ***Title***

  Automatic Construction of Synthetic Musical Instruments and Performers

- ***Meaning***
  - **Framework that builds music synthesis**
  - **Automatic construction**
  - **Instrument model**
  - **Performance model**

Thesis Statement

- **To create a system framework that can automatically create high-quality musical instrument synthesis by using machine-learning techniques to construct the instrument model and the performance model by learning from the performance examples (acoustic recordings and their corresponding scores).**

Start with Modeling…

- **Musical Instrument:**
  - **Trumpet**
- **Music Style:**
  - **Classical Music**

Reasons for Modeling the Trumpet (1)

- **Most wind instrument synthesizers do not sound realistic**
  - **Conflict between:**
    - **Working mechanisms of wind instruments**
      - **Driven by continuous energy exerted by player**
      - **Continuous control drives sound production**
    - **Basic structure of synthesizers**
      - **Mostly sampling-based**
      - **Based on single, isolated notes**
      - **Do not offer a wide range of control**

Reasons for Modeling the Trumpet (2)

- **Previous research**
  - **By Dannenberg and Derenyi (1998)**
  - **Similar scheme**
  - **Produces convincing trumpet sound**

Reasons for Modeling Classical Music

- Characteristics of classical music
  - "Purer" playing style
  - More faithful to the score
  - Fewer articulation effects
- Characteristics of non-classical music
  - Significant inharmonic & transient sounds
  - Need to model noise with a residual model

Contributions of the thesis

- **Use machine learning techniques**
- **Automatically create high-quality synthesis**
- **The problem of control**
  - **Problems of note-oriented synthesis**
  - **Problems of physical models**
  - **This approach simplifies the problem**

Criteria for success

- **Minimum requirement**
  - **Design, implement & test basic framework**
  - **Being able to synthesize realistic trumpet performance for classical music**
  - **Automated modeling process**
  - **Tested on one or two wind instruments.**
- **Extra tasks**
  - **Extend system framework**
    - **Model different musical instruments**
    - **Model different music styles**
- **Future work**

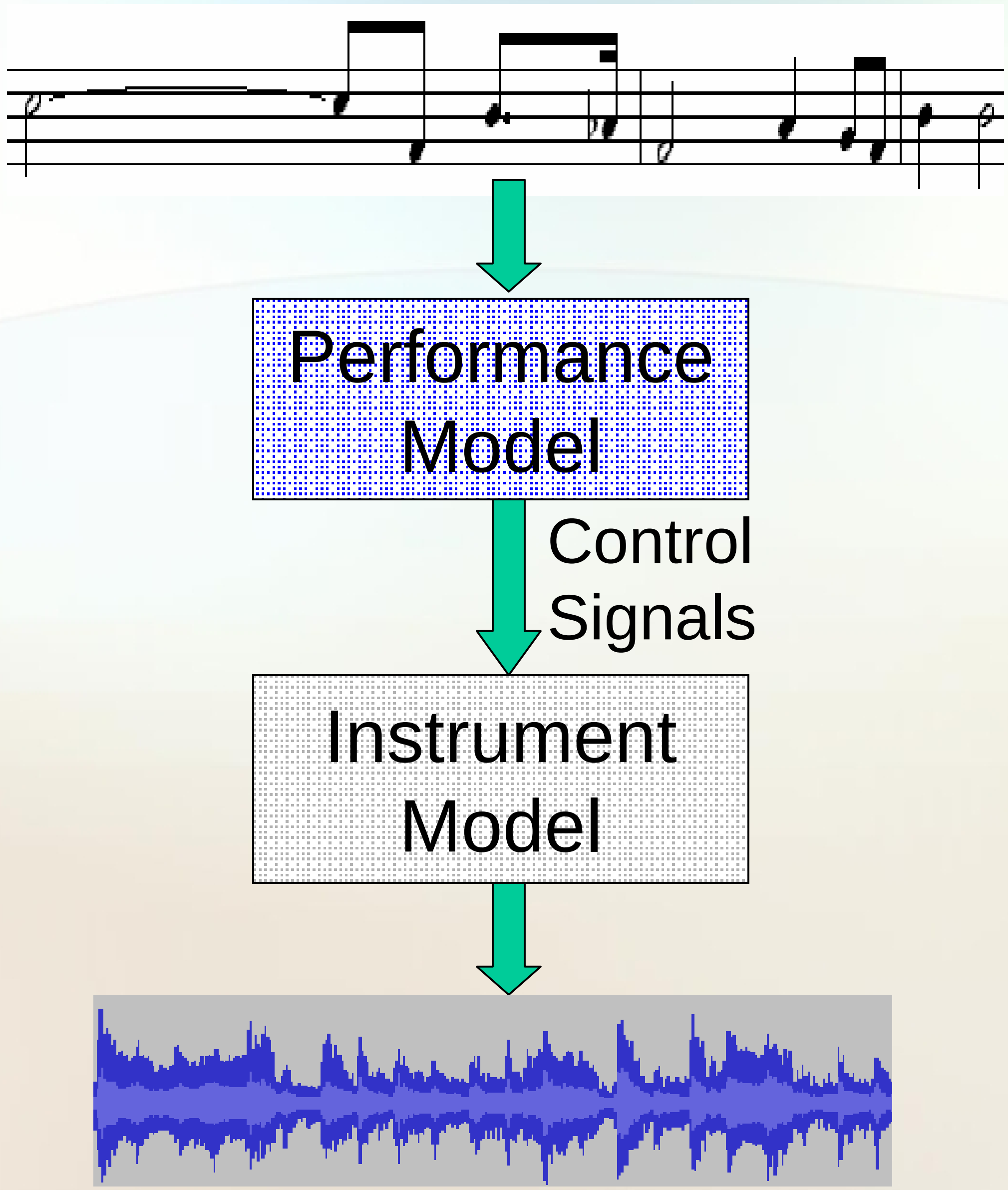
Roadmap

- Introduction
- System Structure
- Main Modules
  - Audio Alignment and Segmentation
  - Instrument Model
  - Performance Model
- Schedule
- Conclusion

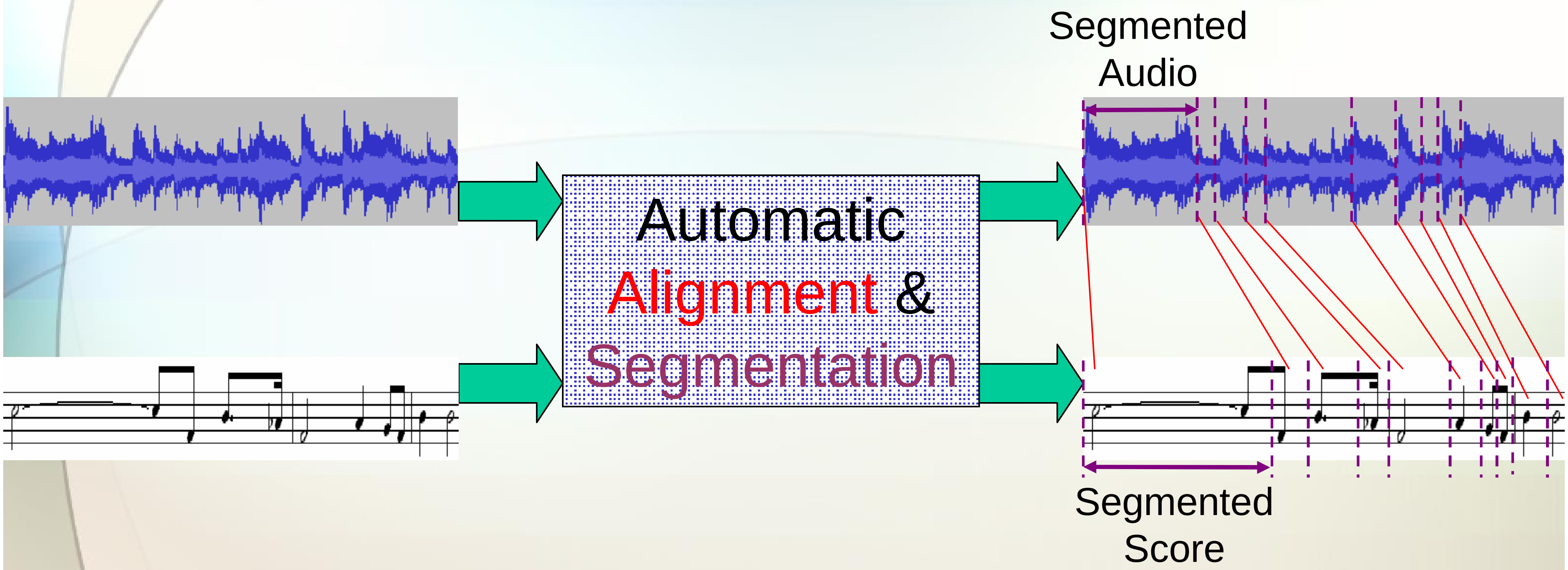
System Structure

- **Synthesis process**
- **Pre-processing training data**
- **Training process for instrument model**
- **Training process for performance model**

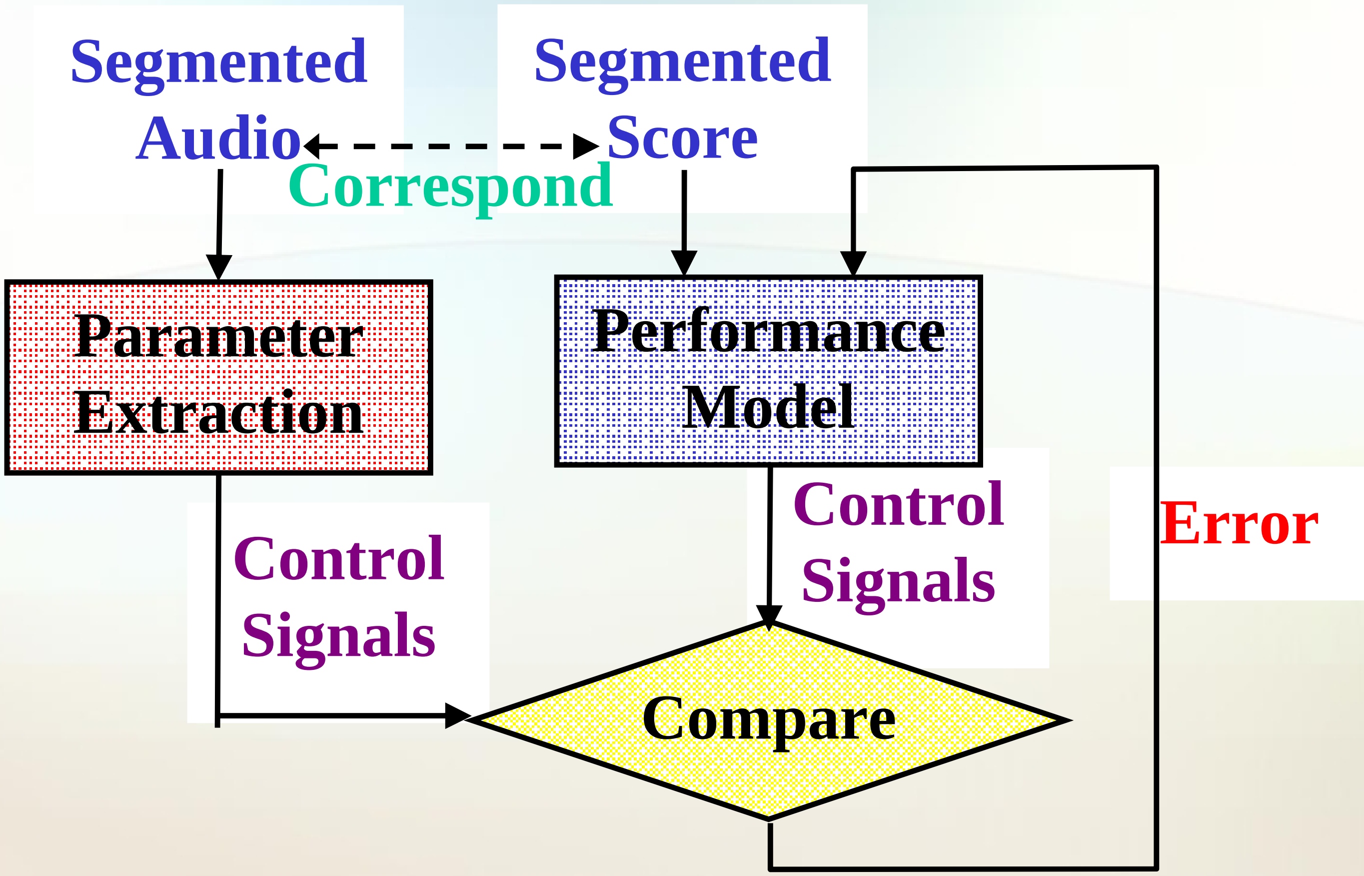
Synthesis Process



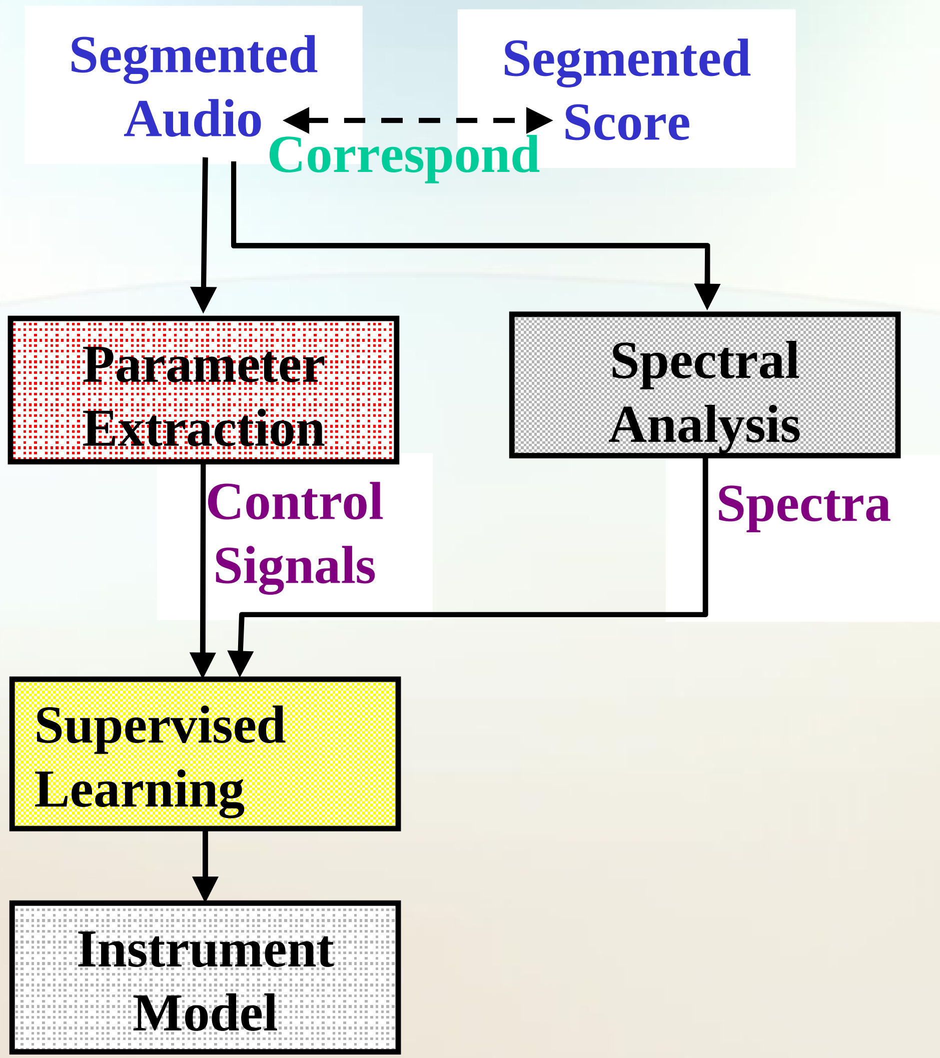
Pre-processing training data



Training the Performance Model



Training the Instrument Model



Roadmap

- ✓ Introduction
- ✓ System Structure
- Main Modules
  - Audio Alignment and Segmentation
  - Instrument Model
  - Performance Model
- Schedule
- Conclusion

Audio Alignment

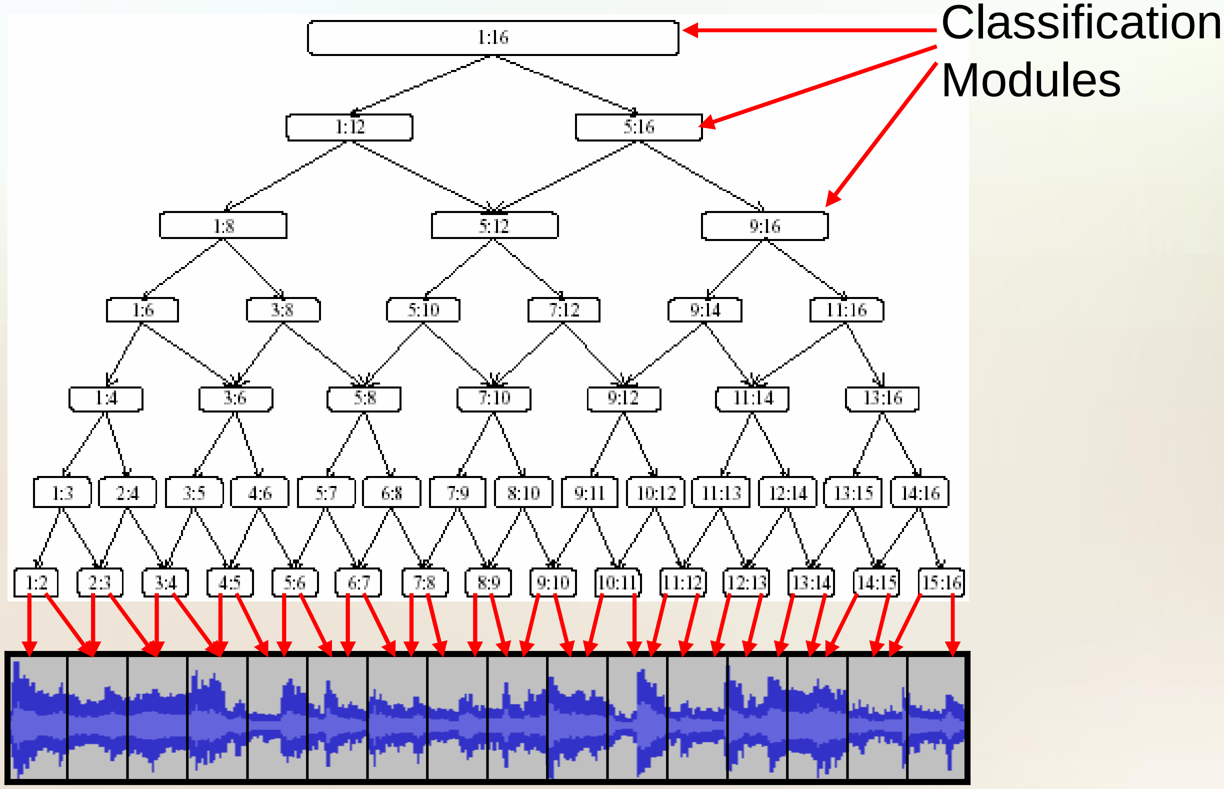
- **Find recording ↔ Correspondence ↔ score**
- **Polyphonic audio alignment**
  - **Extract feature sequences from …**
    - **Acoustic recording**
    - **Score**
  - **Find optimal alignment**
  - **Dynamic programming (DP) or Hidden Markov Model (HMM)**
  - **Satisfactory results**

Audio Segmentation (1)

- **Dannenberg, et. al., (1999) early work**
  - **Define rules & thresholds**
  - **Use features: power; #peaks/period; #zero-crossings/period**
  - **Not reliable and accurate enough**
- **Precise alignment = reliable segmentation**
  - **Require higher accuracy**
  - **Need further modification**

Audio Segmentation (2)

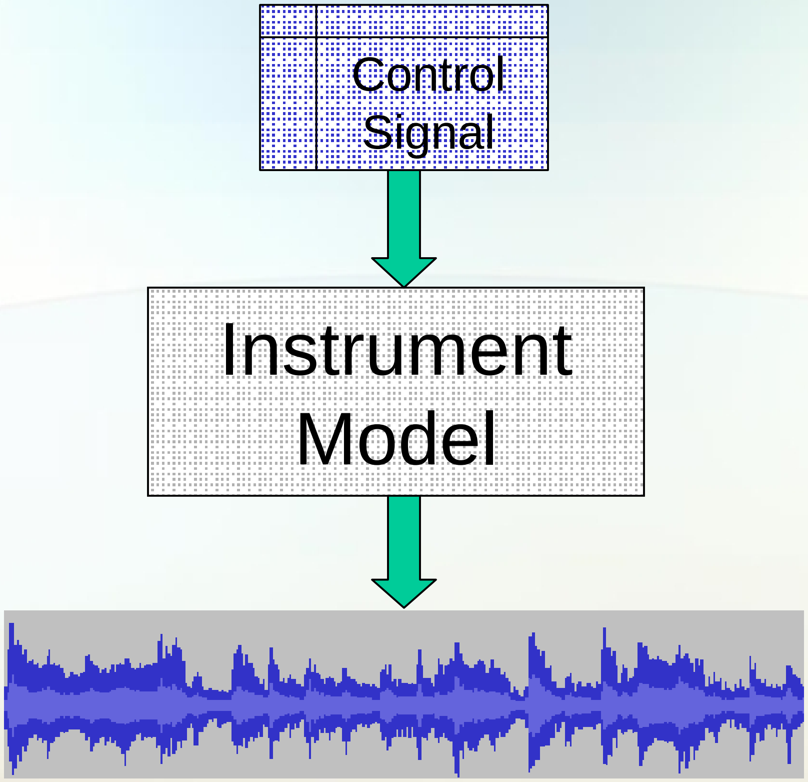
- **Kapanci & Pfeffer's (2004) work**
  - **Segmentation problem ⇨ Classification problem**
  - **Hierarchical machine-learning framework**
  - **Detect soft onset: compare frames separated by increasingly longer distances**



Roadmap

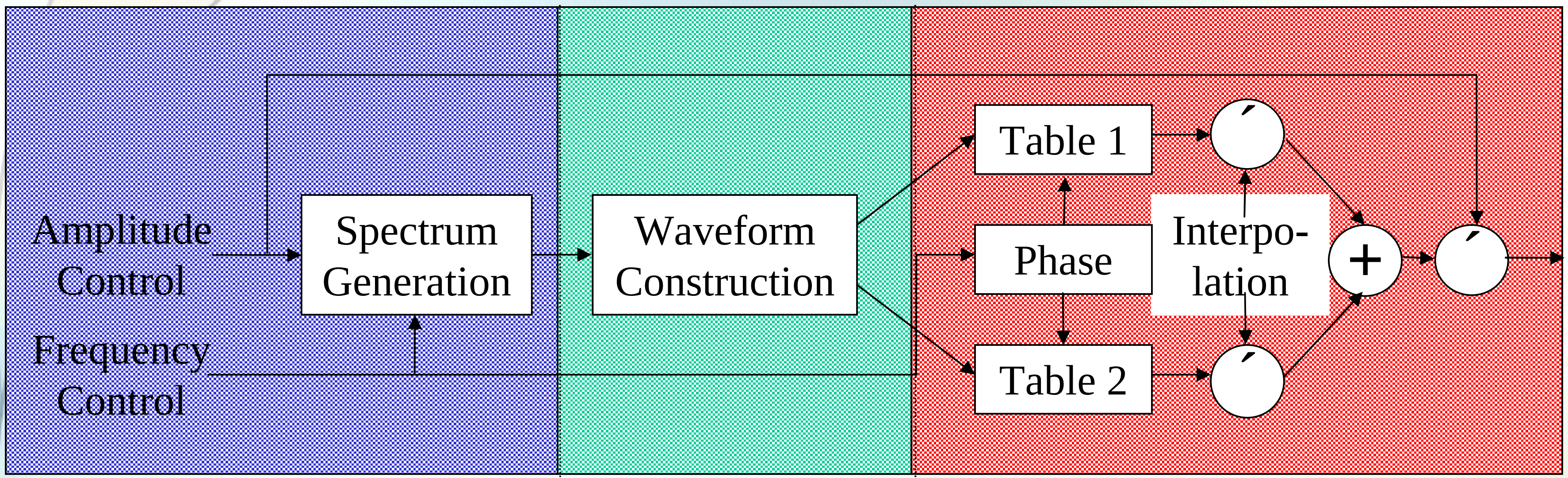
- ✓ Introduction
- ✓ System Structure
- Main Modules
  - ✓ Audio Alignment and Segmentation
  - Instrument Model
  - Performance Model
- Schedule
- Conclusion

Instrument Model



- **Harmonic Model + Residual model**

Harmonic Model



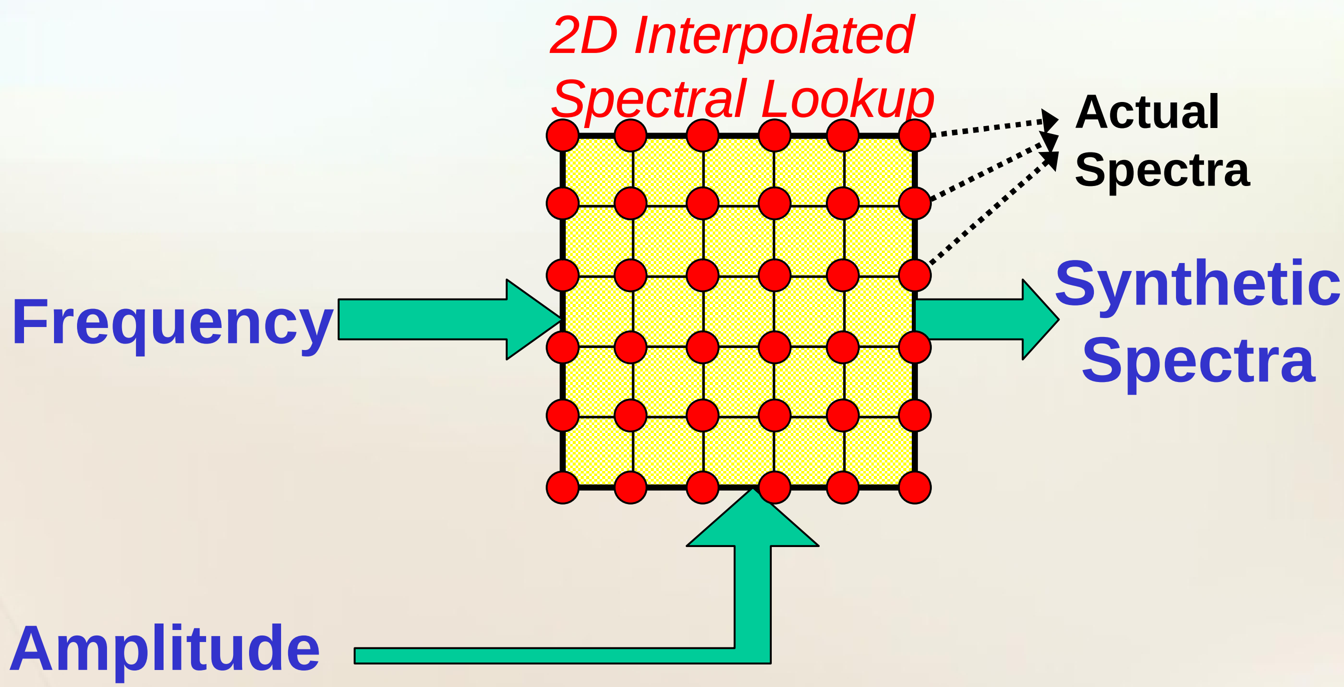
- **Control signals → Spectra**
- **Spectra → Wavetables**
- **Wavetables → Sound samples**

Control Signals → Spectra

- **Spectral interpolation** **(Dannenberg, et. al., 1998)**
- **Memory-based approach** **(Wessel, et. al., 1998)**
- **Neural network** **(Wessel, et. al., 1998)**

Spectral interpolation (Dannenberg, et. al., 1998)

- **Generate Spectral lookup table**
  - **Record a set of sounds**
  - **Obtain a spectrogram for each sound**
  - **Retain specific spectra at thresholds**

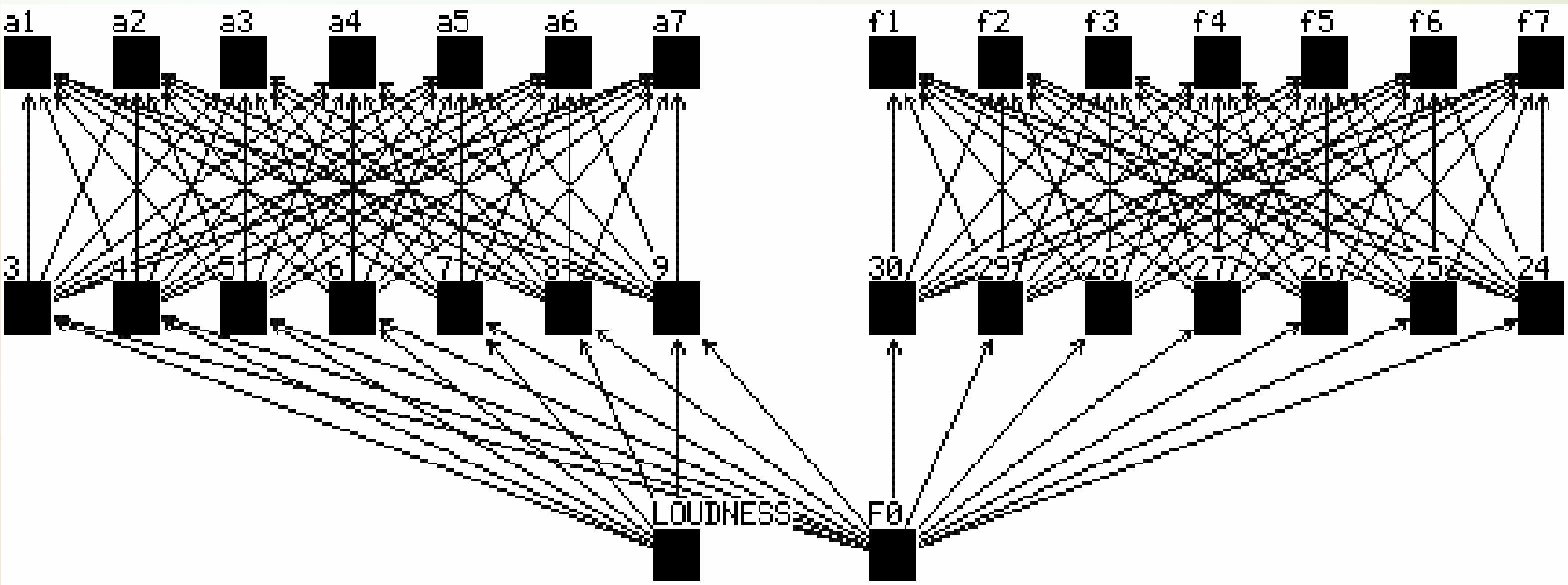


Memory-based approach (Wessel, et. al., 1998)

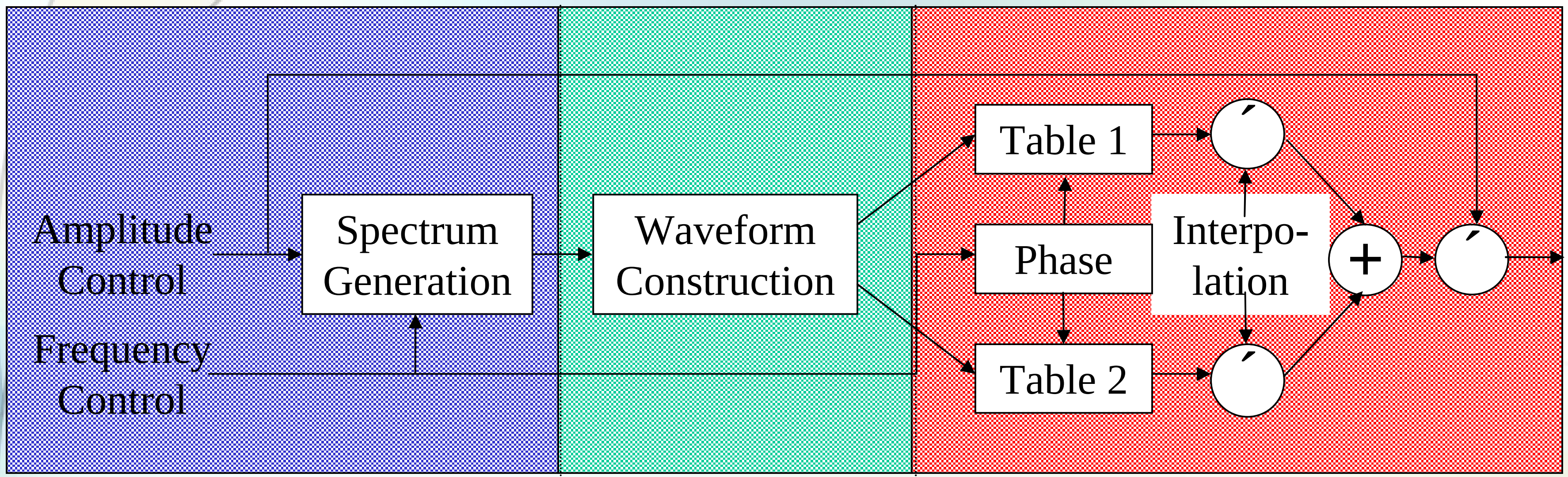
- **Index spectra in a *n* dim. space**
- **Interpolate among *k* nearest neighbors**
- **Special case**
  - **Spectral interpolation technique**
    - ***n*=2 (frequency & amplitude)**
    - ***k*=4**
    - **Linear interpolation**

Neural network (Wessel, et. al., 1998)

- A feed-forward neural network with multiples layers
- Input: frequency, amplitude
- Output: info of sinusoidal components
- Back-propagation learning method
- Advantage: very compact & generalize well



Harmonic Model



- Control signals ➡ Spectra
- Spectra ➡ Wavetables
- Wavetables ➡ Sound samples

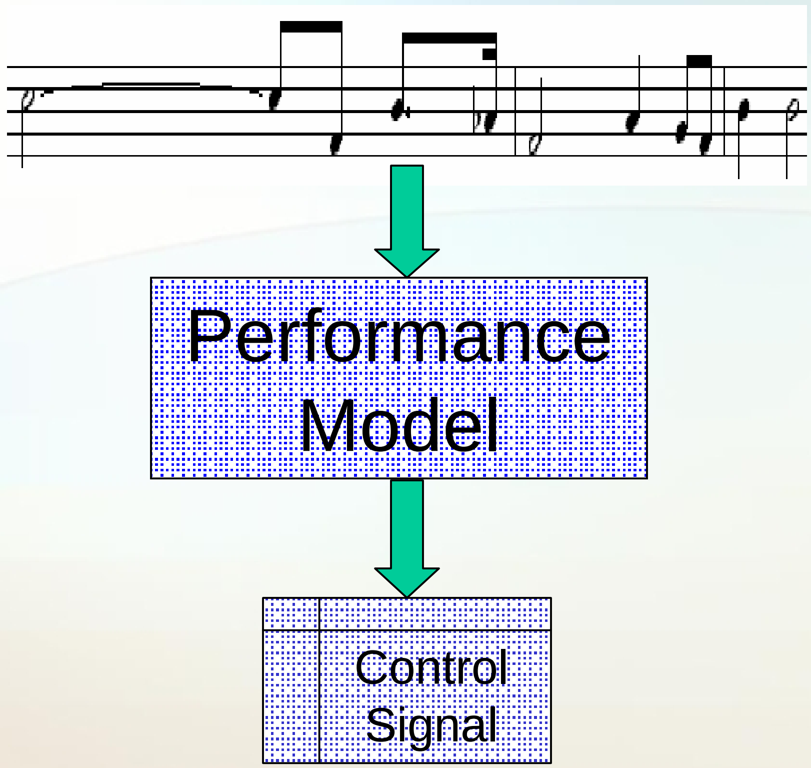
Residual Model

- **Modeling attacks**
  - **Use recorded attacks**
  - **Phase matching**
- **For other instruments**

Roadmap

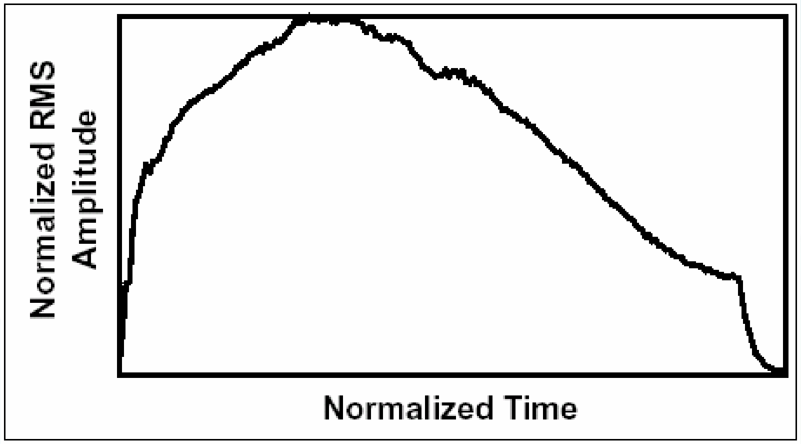
- Introduction
- System Structure
- Main Modules
  - Audio Alignment and Segmentation
  - Instrument Model
  - Performance Model
- Schedule
- Conclusion

Performance Model

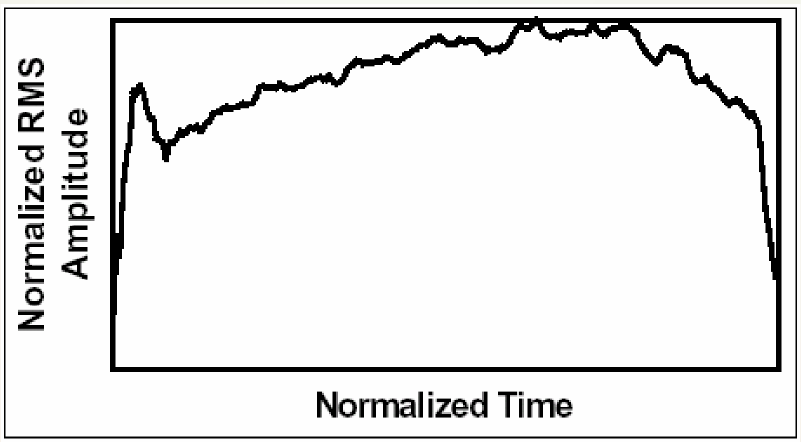


- Control signals
  - Amplitude envelope
  - Frequency envelope
- Error metrics
- Envelope representation
- Mapping scheme

Amplitude Envelope



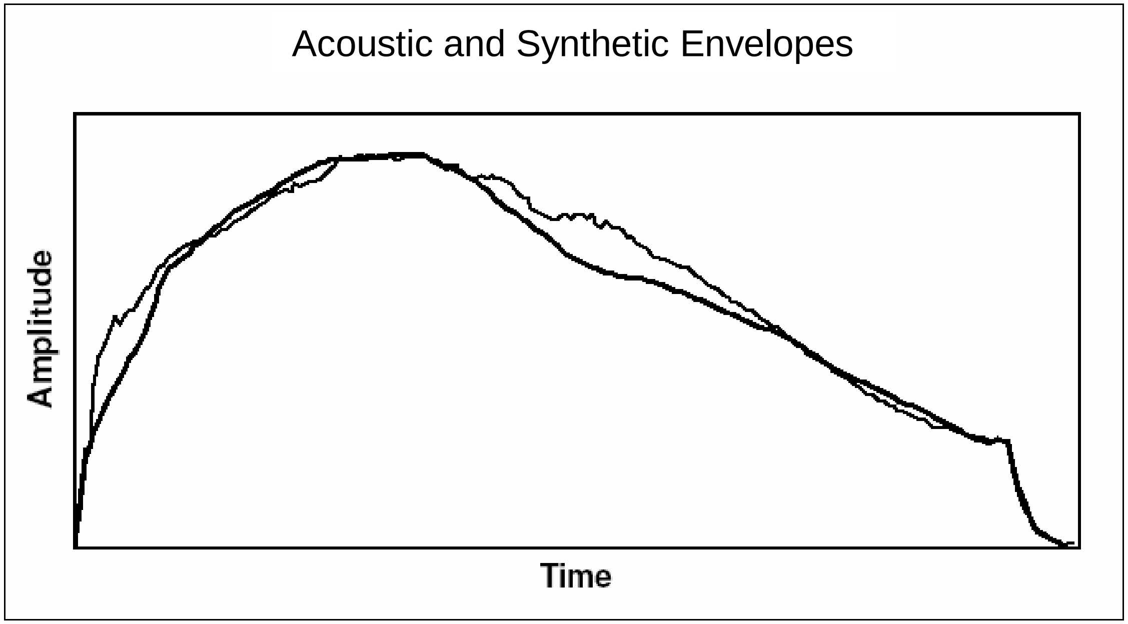
- A tongued note



- A slurred note

*Figures borrowed from (Dannenberg, et. al., 1998)

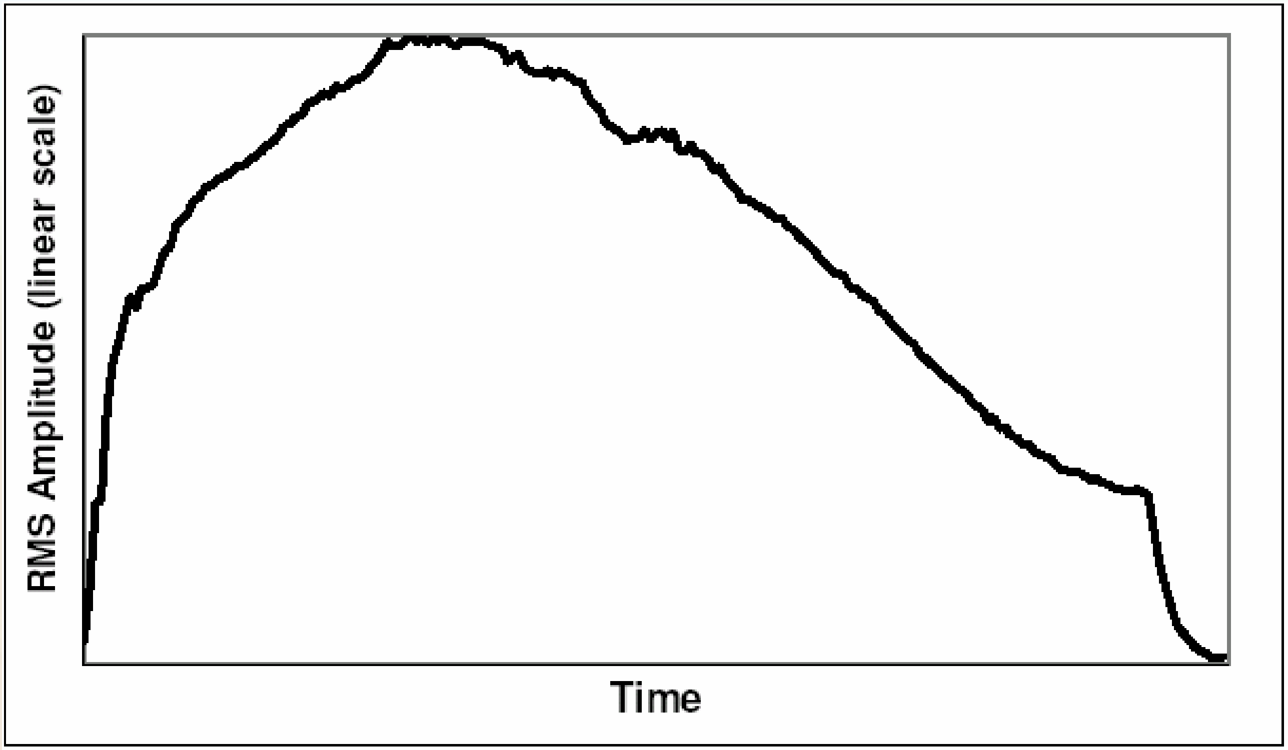
Error Metrics



- **Typical metric: RMS error**
- **Recent work by (Horner et. al., 2004)**
  - **Measure perceptual difference**
  - **Useful reference for error metrics**

Envelope Representation

- **Envelope representation**
  - **Characteristics of amplitude envelope**
    - **Specific duration**
    - **Specific shape**
    - **Specific properties of each part**



Ways to Represent Envelopes

- **Collection of general parameters**
  - **Candidates: center of mass, global/local maximum/minimum, etc.**
  - **Manual vs. automatic selection**
- **Wavelets**
  - **Hierarchical decomposing functions**
  - **Very powerful and popular**

Mapping Scheme

- **Non-linear regression**
  - **Find music context ⟺ actual envelopes**
  - **Examples:**
    - **Neural network**
    - **Kalman filters** $X_k = AX_{k-1} + W_{k-1}$
    - **Function approximation**
- **Pattern clustering**
  - **Classify envelopes into clusters**
  - **For each input data point:**
    - **Use corresponding representative envelope**
    - **Stretch, scale & interpolate accordingly**
  - **Considered as a form of case-based reasoning**

Roadmap

- Introduction
- System Structure
- Main Modules
  - Audio Alignment and Segmentation
  - Instrument Model
  - Performance Model
- Schedule
- Conclusion

Schedule (1)

Oct 2004 ~ Nov 2004 (past) Thesis Proposal	•Collect performance examples for initial experiments •Get familiar with SNDAN package •Read the original code by Dannenberg & Derenyi •Propose thesis topic
Dec 2004 ~ May 2005	**System Development**
Dec 2004 ~ Jan 2005	Implement Audio Alignment & Segmentation module
Jan 2005	Incorporate SNDAN package to training data pre-processing stage
Feb 2005	Develop and compare instrument models
Mar 2005 ~ Apr 2005	Design and implement performance model
May 2005	System integration and testing

Schedule (2)

Jun ~ Aug 2005	**System Evaluation & Tuning**
Jun 2005	Model other instruments
Jul 2005	Synthesize other types of music
Aug 2005	System and model evaluation and fine tuning
Sep ~ Nov 2005	**Writing Thesis**
Sep 2005 ~ Nov 2005	Thesis write-up and revisions
Nov 2005	Thesis defense

Conclusion

- **Propose a scheme:**
  - **automatically construct**
  - **high-quality instrument synthesis**
  - **by learning from performance examples**
- **Machine learning ➡ a crucial role.**
- **Future work**
  - **Modeling different instruments**
  - **Modeling different music styles**
  - **Make it work in real-time**

Roadmap

- ✓ **Introduction**
- ✓ **System Structure**
- **Main Modules**
  - ✓ **Audio Alignment and Segmentation**
  - ✓ **Instrument Model**
  - ✓ **Performance Model**
- ✓ **Schedule**
- ✓ **Conclusion**

Thank You

- **Acknowledgements**
  - **Roger B. Dannenberg**
  - **Istvan Derenyi**
  - **James Beauchamp (SNDAN)**
- **Contact**

Ning Hu (ninghu@cs.cmu.edu)