Toward Automatic Transcription
-- Pitch Tracking In Polyphonic Environment

Term Project Presentation By:
Keerthi C Nagaraj
Dated: 30th April 2003
Outline

- Introduction
- Background
  - problems in polyphonic pitch tracking
- Previous approaches
  - Sinusoidal Modeling
  - Auditory Modeling
- Current Approach
  - Use of Prior Knowledge - Bayesian Probability Network
  - Implementation
- Results
- Conclusion
Introduction

What do we have?

What do we need?

Keerthi C Nagaraj, Department of Electrical & Computer Engineering
Pitch estimation Process:

1. Segmentation /Rhythm tracking
2. "pitch info" extraction
3. Most probable F₀ Candidates
4. Eliminate interfering harmonics
5. Best pitch estimate

Feature analysis
Tone model

Keerthi C Nagaraj, Department of Electrical & Computer Engineering
Problems with Polyphonic Pitch extraction

- Mathematically ambiguous problem
  - Overlapping partials
  - Expressionist performance, not traceable
  - Onset asynchronies
  - Percussion sounds in “real world” signals

Keerthi C Nagaraj, Department of Electrical & Computer Engineering
Past work

**Sinusoidal Model:**

STFT, Constant Q transforms, Bounded Q transforms

- More focussed on forming a mathematical model of pitch perception

**Auditory Model:**

- Lyon’s Cochlear Model, Meddis & Hewitt Model
- More focussed on laying a perceptual background

Keerthi C Nagaraj, Department of Electrical & Computer Engineering
Encountered Problems

- They do not eliminate the confusion due to overlapping partials
- Frame to Frame independent calculation

Approach:

- Use higher level knowledge
- Cross frame data integration
- Probabilistic/ ‘belief based’ approach
Current approach

- **Step 1:**
  - Using Auditory model to extract sound as perceived by the ear
Current Approach (Contd.)

- Step 2:
- Extract pertinent features of the sound (Loudness, $F_0$ & color)--use of Summary Auto-Correlation Function (SACF)
Bayesian modeling

Step 3: use of the features as knowledge base

A simple graphical model with observed data $d$, patterns $\{a, b\}$ and hyperparameters $\{c, e\}$.

Each $\bigcirc$ stands for an individual perceptual sound.
Implementation

Assign a priori pdfs to the parameters =>

The joint posterior probabilities are obtained as:

\[
p(\omega_i^q | \Delta_{\omega}, \delta_{\omega}^2) = N(\Delta_{\omega}, \delta_{\omega}^2)
\]

\[
p(H_i^q) = \mathbb{I}_{[1,H_{\text{max}}]} \frac{1}{H_{\text{max}}}
\]

\[
p(b_i^q | \omega_i^q, H_i^q, \sigma_i^2, \delta_{\omega}^2) = N(0, \sigma_i^2 \Sigma(\omega_i^q, H_i^q, \delta_{\omega}^2))
\]

where \( \Sigma(\omega, H, \delta^2) = \delta^2 (G^t G)^{-1} \)

\[
p(\Gamma^q) = \alpha \Gamma^q (1 - \alpha) (1 - \Gamma^q)
\]

\[
p(\sigma_i^2) = IG(\alpha_\sigma, \beta_\sigma)
\]

\[
p(\Delta_{\omega}) = N(\Delta_{\omega}^{\text{prev}}, \sigma_{\Delta}^2)
\]

\[
p(M_Q, \{ \hat{\theta}^q \}_Q | d, \delta^2) \propto \frac{p(\{ \hat{\theta}^q \}_Q | M_Q) p(M_Q)}{(1 + \delta^2)^{\frac{M}{2}}} [d' P_c d + 2\beta_c]^{-\epsilon}
\]

\[
P_c = I_N - \frac{\delta^2}{1 + \delta^2} G_c (G_c^t G_c)^{-1} G_c^t
\]

Where \( M = 2\Sigma_{q \in Q} H_q, \theta^q = \{ \omega^q, H^q \}, \epsilon = N/2 + \alpha, \quad p(M_Q) = p(\{ \Gamma^q \}_{q = 1:Q}) \)

\( \sigma^2 \) represents the expected SNR, \( G_c \) Composite basis matrix

Reference: Wamsley Godsill & Rayner -1999

Keerthi C Nagaraj, Department of Electrical & Computer Engineering
For each multi-frame,
- Collect the peaks, multiply with the reliability vector
- pass the output through a weighted median filter
- Find error by comparing the evolving model and the observed data
- Update the reliability vector
- repeat the process to minimize the error
Results
Results (Contd.)

![Graphs showing results](image)
Conclusion & Future work

- Auditory model for pitch perception was implemented
- Hierarchy of music information was modeled as a simple Bayesian probability network
- Pitch tracking was done using auditory model front end processing and knowledge based resolving of partials
- Beat tracking can be done to shorten focus of pitch detection to the steady state areas of sound
- Other auditory “cues” can be added to the BPN.
- Musical Instrument models can be used to enhance the transcription process
- Feasibility of adding new parameters can be tested for impact on transcription.