SINUSOIDAL MODELING

EE6641 Analysis and Synthesis of Audio Signals
Yi-Wen Liu
Nov 3, 2015
Last time: Spectral Estimation

• Resolution
  • Scenario: multiple peaks in the spectrum
  • Choice of window type and window length
  • Rule of thumb: $\Delta \omega = O(1/\Delta \tau)$

• Accuracy
  • Scenario: signal in noise
  • Quadratic interpolation of FFT log-magnitude
  • Fisher Information and Cramer-Rao lower bounds
  • Coherent frequency estimation: $\Delta \omega = O(\Delta \tau^{-3/2})$
Today’s agenda

• Sinusoidal modeling
  • Analysis
    • Formula for QI-FFT
    • Choice of window
    • Peak tracking
  • Synthesis
  • Applications

• Spectral decomposition
  • (S + N) vs. (S + N + T)
  • Spectral subtraction
QI-FFT: a fast frequency estimator

- **Windowing**: time-domain multiplication with a shaping function
- **Zero-padding**: append zeros in time $\Leftrightarrow$ interpolate in frequency
- **FFT**: apply fast Fourier transform
- **Peak detection**: find every peak
- **Quadratic interpolation**: fit a parabola in log-magnitude

$\Rightarrow$ The location of the parabola is the frequency estimate.

A formula for QI-FFT

\[ L(x) = Ax^2 + BX + C \]
\[ L(-1) = L^- \]
\[ L(0) = L^0 \]
\[ L(1) = L^+ \]

\[ A = \frac{L^+ + L^- - 2L^0}{2} \]
\[ B = \frac{(L^+ - L^-)}{2} \]
\[ C = L^0 \]

\[ L(x) = A(x - q)^2 + \hat{L} \]
\[ q = -\frac{B}{2A} \]
\[ \hat{L} = L^0 - \frac{B^2}{4A} \]
Example: finding peaks in a spectrum

- Peaks can move
- Peaks can grow or attenuate
- Need to ignore side lobes
  - But how?
Blackman window has > 50 dB side-lobe suppression

- All side lobes beneath the *masking threshold.*
Example:
Hann-windowed transform of trumpet sounds

signal spectrum

Masked region
Psychoacoustic masking

- **Masking** refers to the fact that softer sounds cannot be heard due to the presence of stronger sounds.
  - Forward masking
  - Simultaneous
  - Backward masking

- We will cover psychoacoustics later in this semester.
Sinusoidal Modeling: from discrete peaks to trajectories
Trajectory formation

• Basic peak tracking involves:
  • Finding shortest link
  • Resolving splits

• Advanced technique includes
  • Forbidding large jumps
  • Transient detection
  • Highest peak chooses first
Frequency-trajectories of a speech signal, arrows indicate *transient regions*
Part II: Synthesis

• Sinusoidal modeling
  • Analysis
  • **Synthesis**
    • Linear-interpolation
    • Phase continuation
    • A window-based synthesis method

• Spectral decomposition
  • \((S + N)\) vs. \((S + N + T)\)
  • Spectral subtraction
  • Noise modeling (Nov. 16, 23)
Sinusoidal synthesis (ii): the linear interpolation method

\[ \{A_m\}_{m=0}^{M-1} \quad \text{.........an amplitude envelope} \]

\[ \{f_m\}_{m=0}^{M-1} \quad \text{.........a frequency envelope} \]

\[ A[n] = \beta A_m + (1 - \beta)A_{m+1} \]

\[ f[n] = \beta f_m + (1 - \beta)f_{m+1} \]

\[ \beta = \frac{(m+1)h-n}{h} \]

Remark: amplitude can be log or linear scale.
The linear-interpolation method (cont’d)

- Between time $mh$ and $(m+1)h$, do this:

  \[ A[n] = \beta A_m + (1 - \beta) A_{m+1} \]
  \[ f[n] = \beta f_m + (1 - \beta) f_{m+1} \]

- update the phase sample-by-sample:

  \[ \varphi_n = \varphi_{n-1} + \frac{2\pi f[n]}{f_s} \]

- For each trajectory $j$, do:

  \[ s_j[n] = A[n] \cos(\varphi_n) \]

- Finally, sum over all $j$
From one trajectory to a signal: a window-based method

\[ \{ A_m \}_{m=0}^{M-1} \quad \text{an amplitude envelope} \]

\[ \{ f_m \}_{m=0}^{M-1} \quad \text{a frequency envelope} \]

Synthesis:

\[ s[n] = \sum_{m=0}^{M-1} A_m w[n - mh] \cos (2\pi f_m n T + \phi_m) \]

Continuous phase updates:

\[ \phi_m = \phi_{m-1} + 2\pi \left( \frac{f_{m-1} + f_m}{2} \right) hT \]
summation over all trajectories

\[ s[n] = \sum_{\text{trajectories}} \sum_{m=0}^{M-1} A_m w[n - mh] \cos(2\pi f_m nT + \phi_m) \]
Applications of sinusoidal modeling

• Noise removal/ speech enhancement
  • Potentials in hearing aids/ cochlear implants

• Musical effects:
  • Pitch shifting/ time warping

• Parametric audio coding:
  • MPEG-4 structured audio
Part III: Spectral decomposition

• Sinusoidal modeling
  • Analysis
  • Synthesis
    • Linear-interpolation
    • Phase continuation
    • A window-based synthesis method

• Signal decomposition
  • Spectral subtraction
  • (S + N) vs. (S + N + T)
  • Noise modeling (Possible final project idea)
Example: a watermarking system based on signal decomposition (Liu & Smith, 2007)
Sines + Noise + Transient (S+N+T) decomposition
Spectral subtraction

- For each peak in the spectrum, do
  1. fit the mainlobe of Blackman window
  2. subtract the mainlobe
  3. Inverse FFT to synthesize the residual (“noise”)
Sines + “noise” decomposition by spectral subtraction

(a) Trumpet

(b) Dance music

Parametric noise modeling?
Transient detection

Audio in → Blackman window → STFT → Peak detection

Peak detection → Peak tracking → Sinusoidal synthesis

I-FFT → OLA → Residual

Energy ratio → SRR

Previous frame

> 1.5? → Transient Y/N

< 0.5?
Transient detection: there’s no gold standard

- The following is just a heuristic approach
  - Energy comparison
  - Sine-to-residual ratio (SRR)
Entering and exiting the transients

- Once a transient is detected, sinusoidal model is halted for a period of time.
Possible project ideas

- Trajectory formation
  - Causal implementation for real-time applications
- Applications of spectral subtraction
  - Noise-removal / Speech enhancement
- Audio recognition
  - Sound source segregation
  - Audio fingerprinting/ watermarking
  - And so on…