From rat laughter to speed metal - a personal history of phase vocoding

George Tzanetakis (gtzan@cs.uvic.ca)

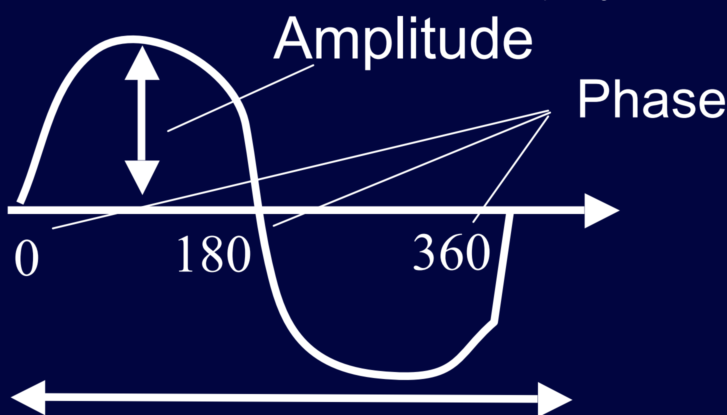
Assistant Professor

Computer Science Department

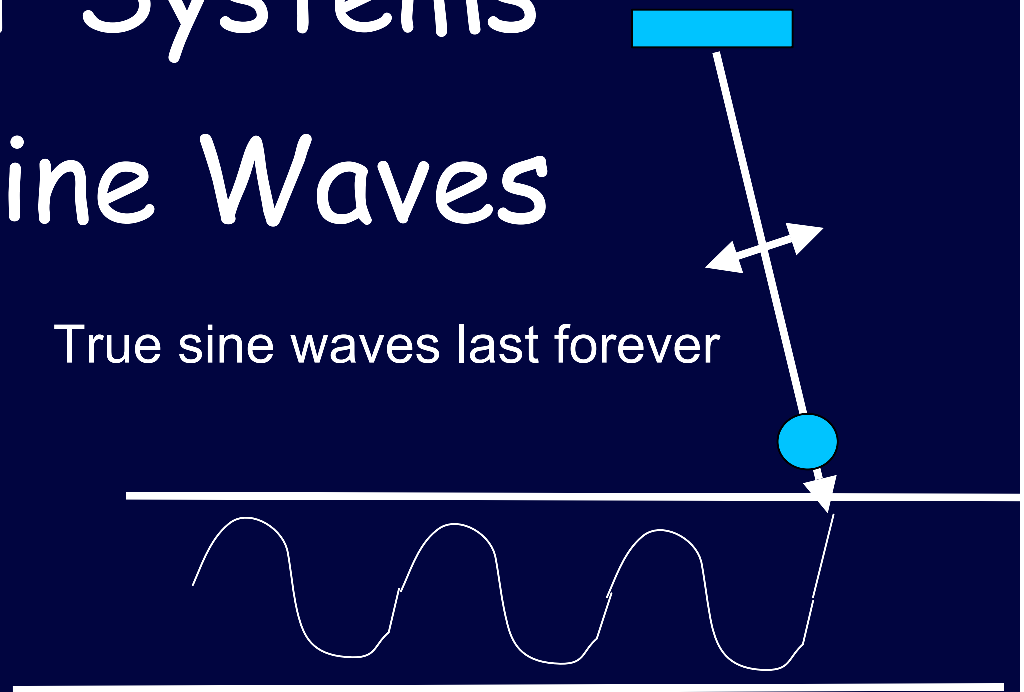
(also in Music, ECE)

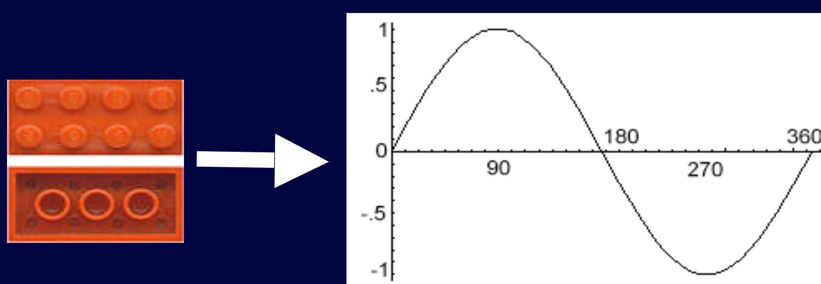
University of Victoria, Canada

Linear Systems and Sine Waves

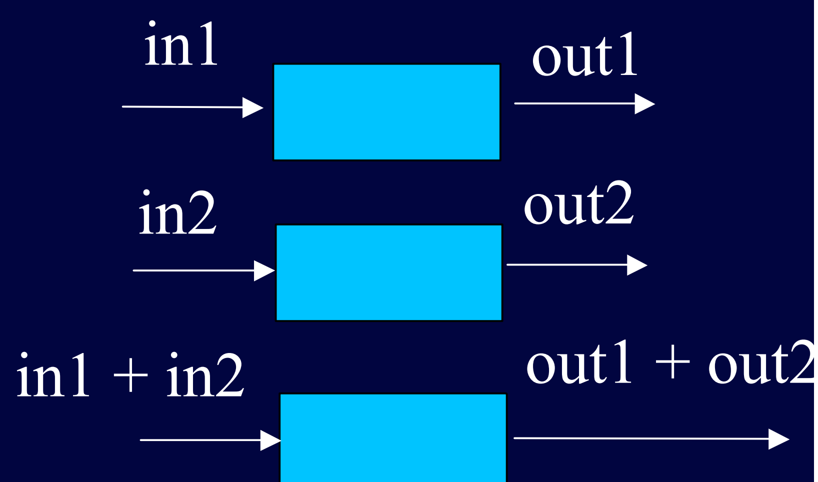


Period = 1 / Frequency





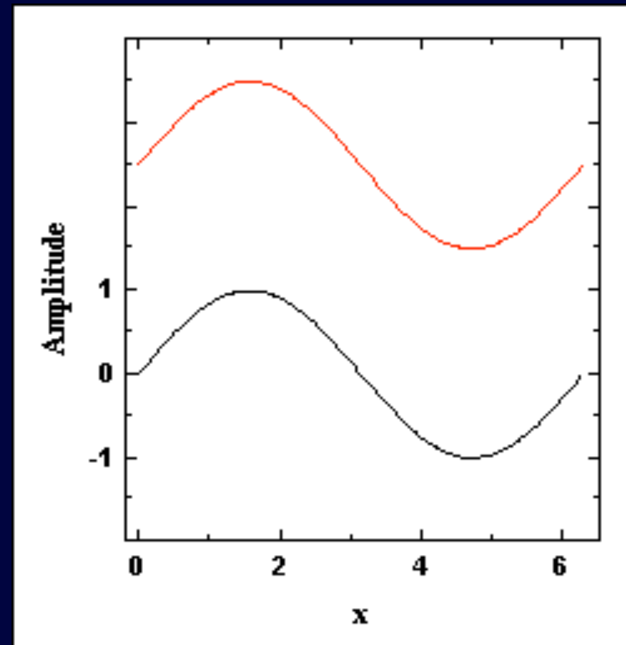
sine wave -> LTI -> new sine wave



Time-Frequency Analysis

Fourier Transform

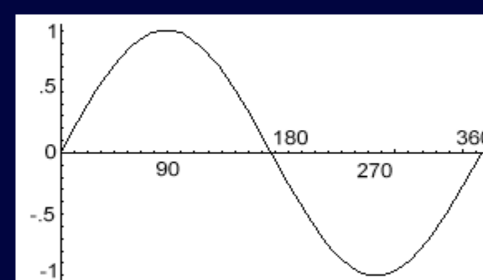
$$f(x)=\sum_{n=0}^{\infty} a_n \cos(n*x)+\sum_{n=0}^{\infty} b_n \sin(n*x)$$



$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} f(\omega)\,e^{-i\omega t}\,dt$$

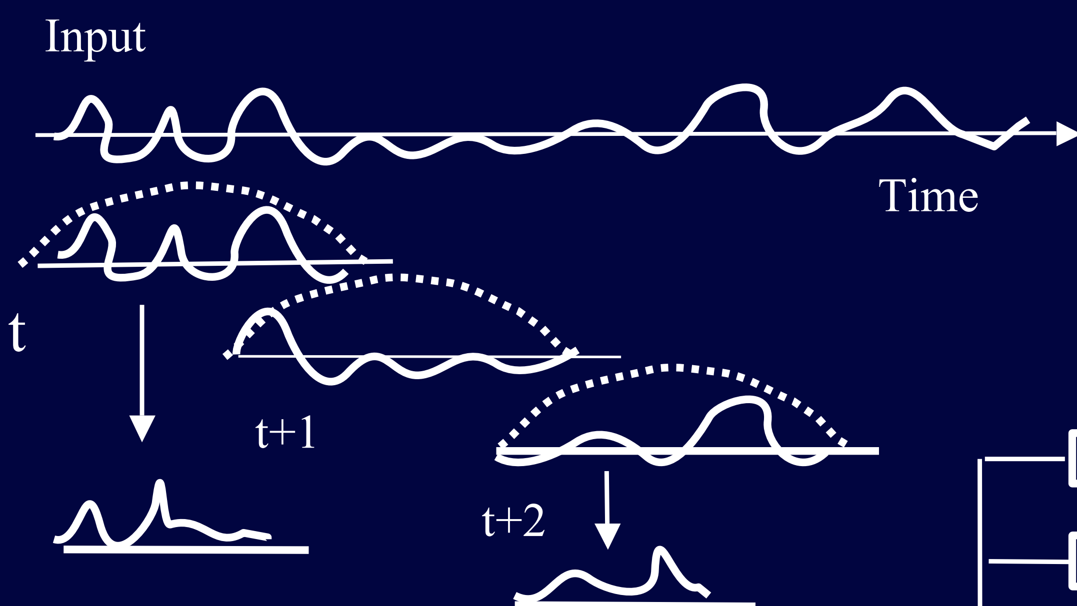
$$f(\omega)=\int_{-\infty}^{\infty} f(t)\,e^{i\omega t}\,dt$$

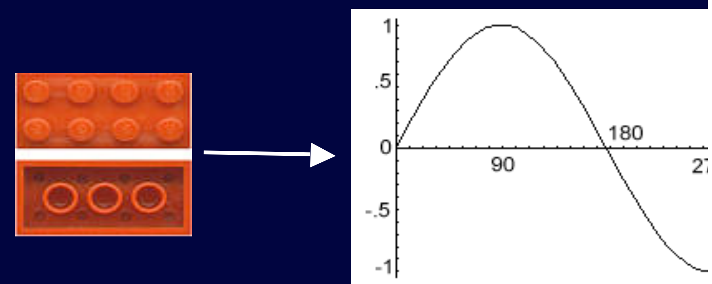
$$e^{i\theta}=\cos(\theta)+i*\sin(\theta)$$

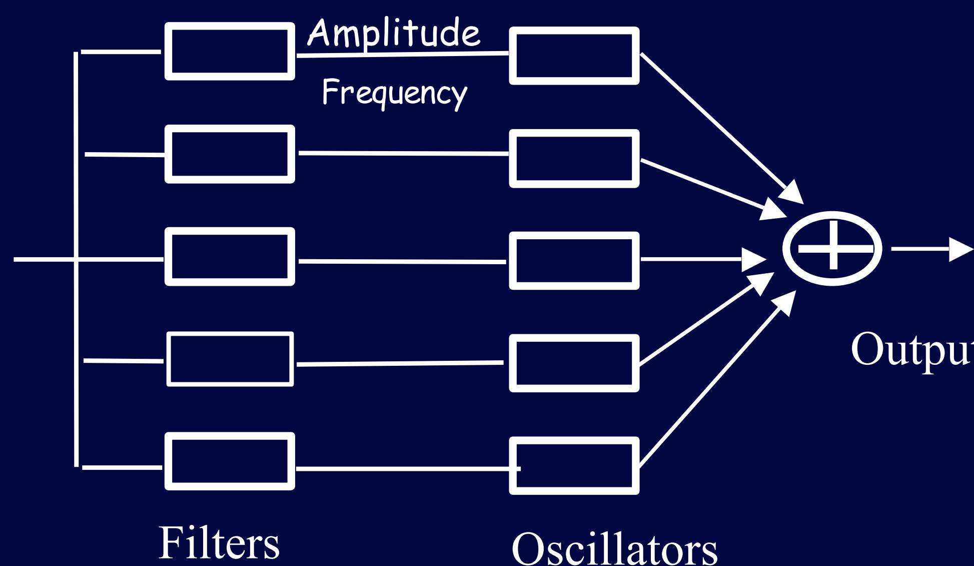


e^{i}— = cos___ *i*_sin___

Short Time Fourier Transform







Time-varying spectra
Fast Fourier Transform FFT

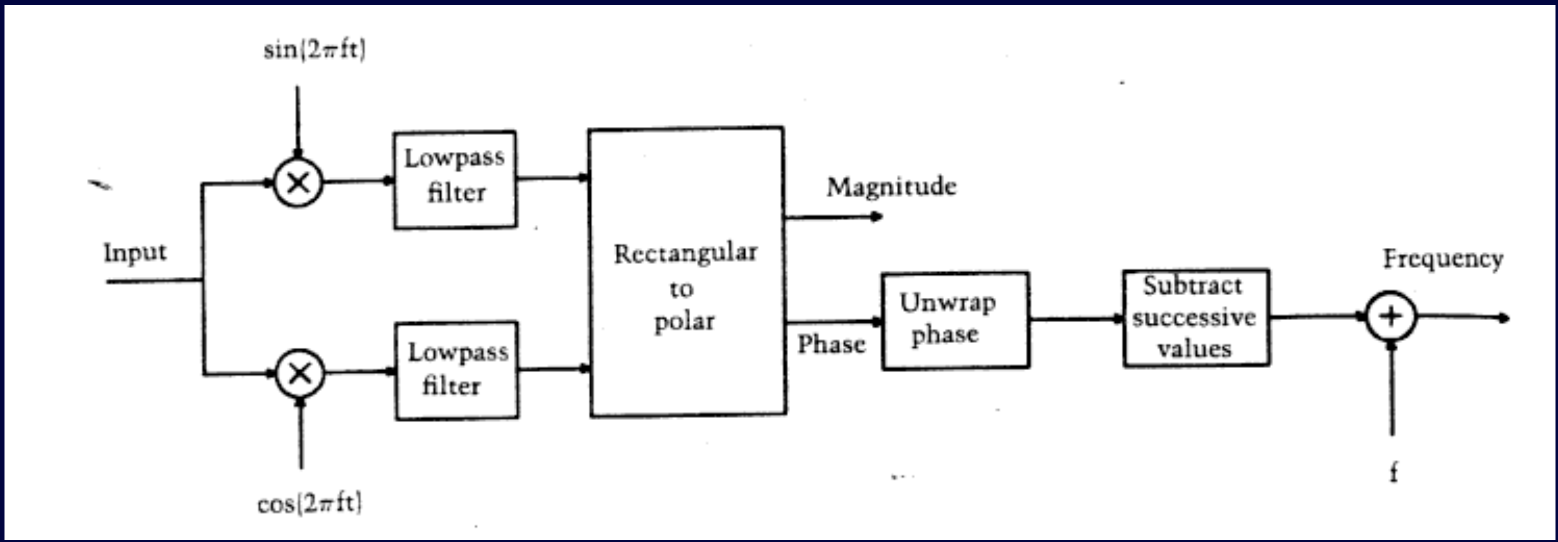
Phasevocoder

- Vocoder = voice encoder (Youtube: Herbie Hancock - I thought it was you)
- Phase vocoder = vocoder that takes into account phase information
- Time stretching and pitch shifting
- Variable speed playback changes the pitch
- Why ?

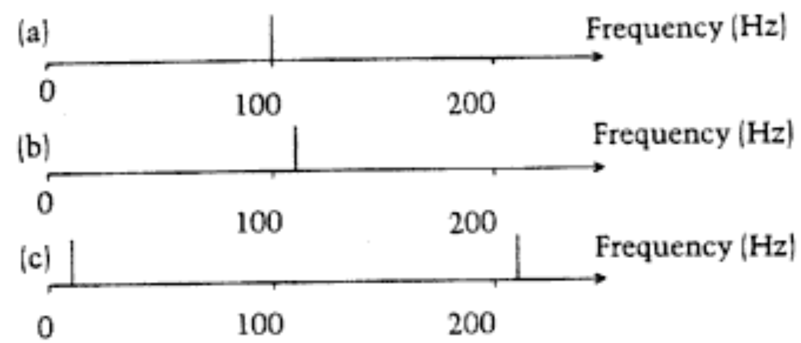
Phase-vocoding

- 1966 Flanagan and Gold (Bell Labs, Speech)
- Analysis/Synthesis system
- Signal = sum of time-varying sine waves (amplitude, frequency and phase)
- Filterbank constraints = freq. repsonse of each filter identical except center frequency, sum has flat freq. response, center frequencies equally spaced

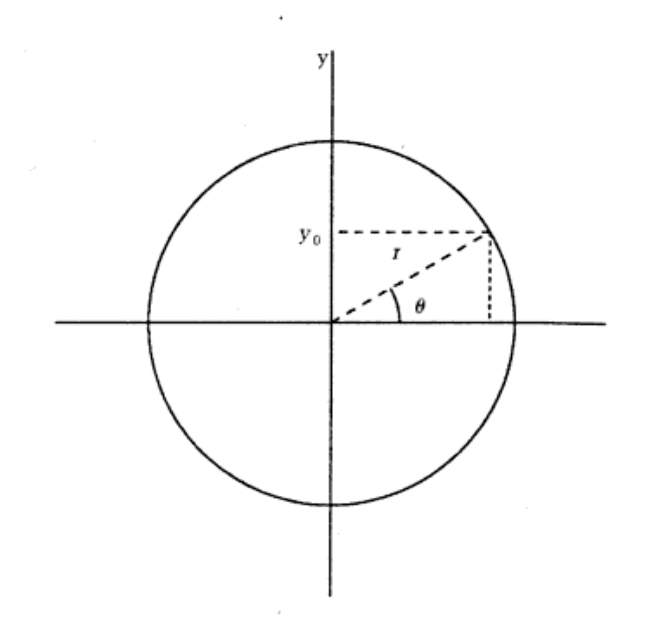
Band-Pass Filter



Heterodyning

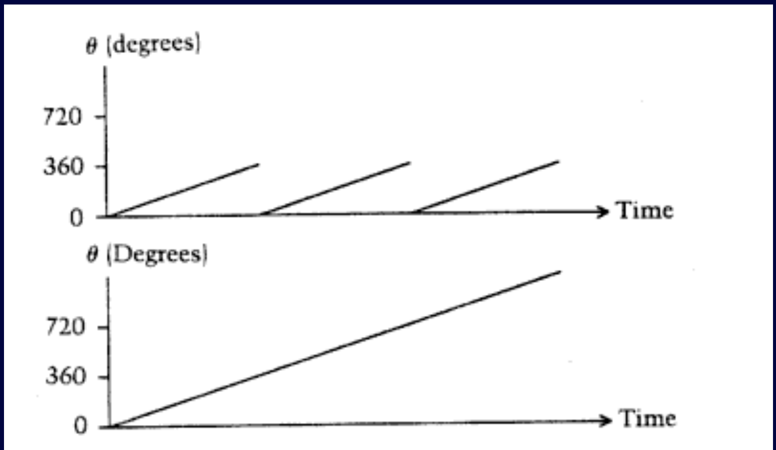


Phasors



$$r = \sqrt{x_0^2 + y_0^2}$$

$$\theta = arctan\left(\frac{y_0}{x_0}\right)$$



Rectangular -> Polar

Time varying amplitude = radi
Time varying frequency = rate
angular rotation

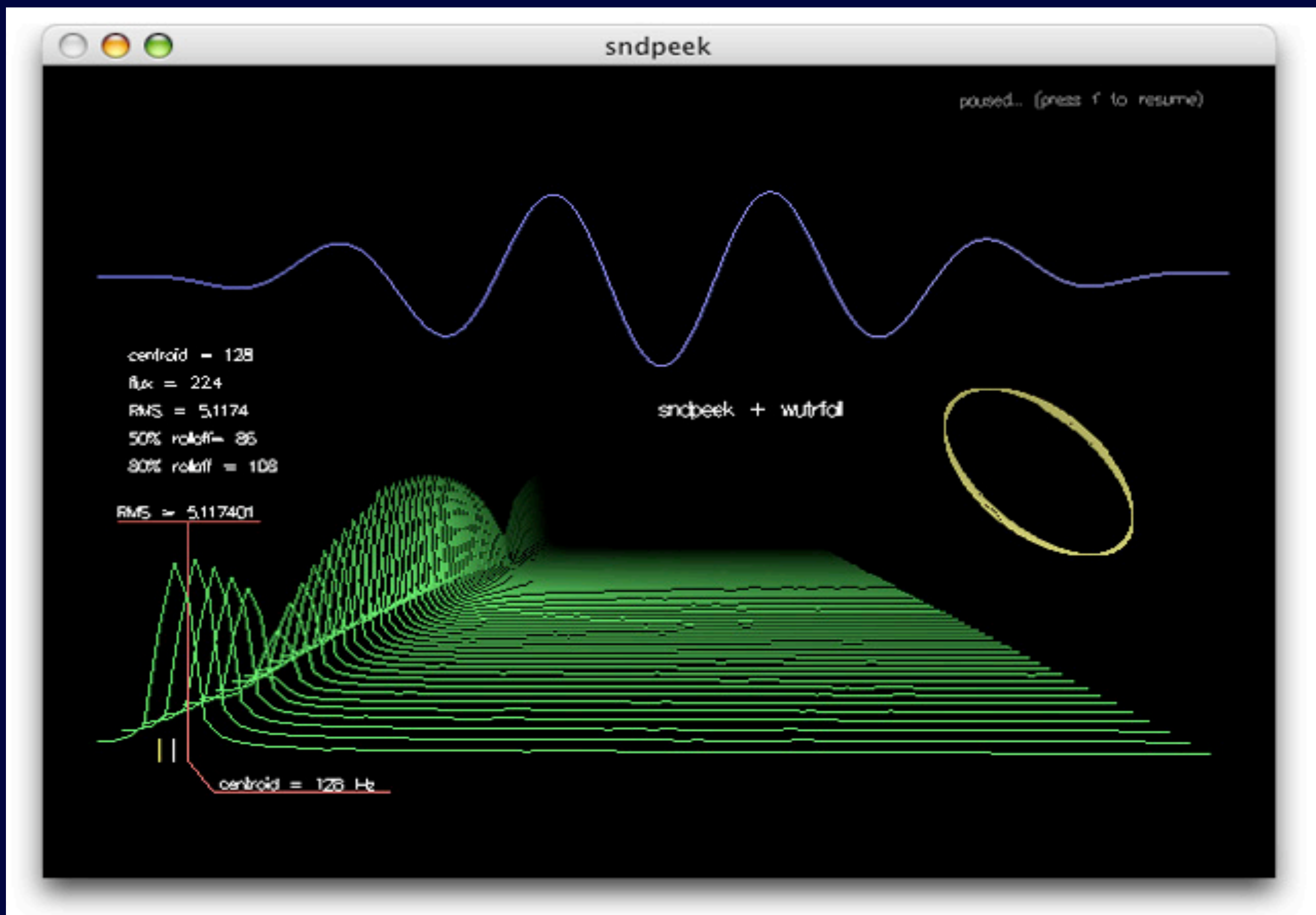
only 0 to 360 degrees

What if we have the followin
sequence ?

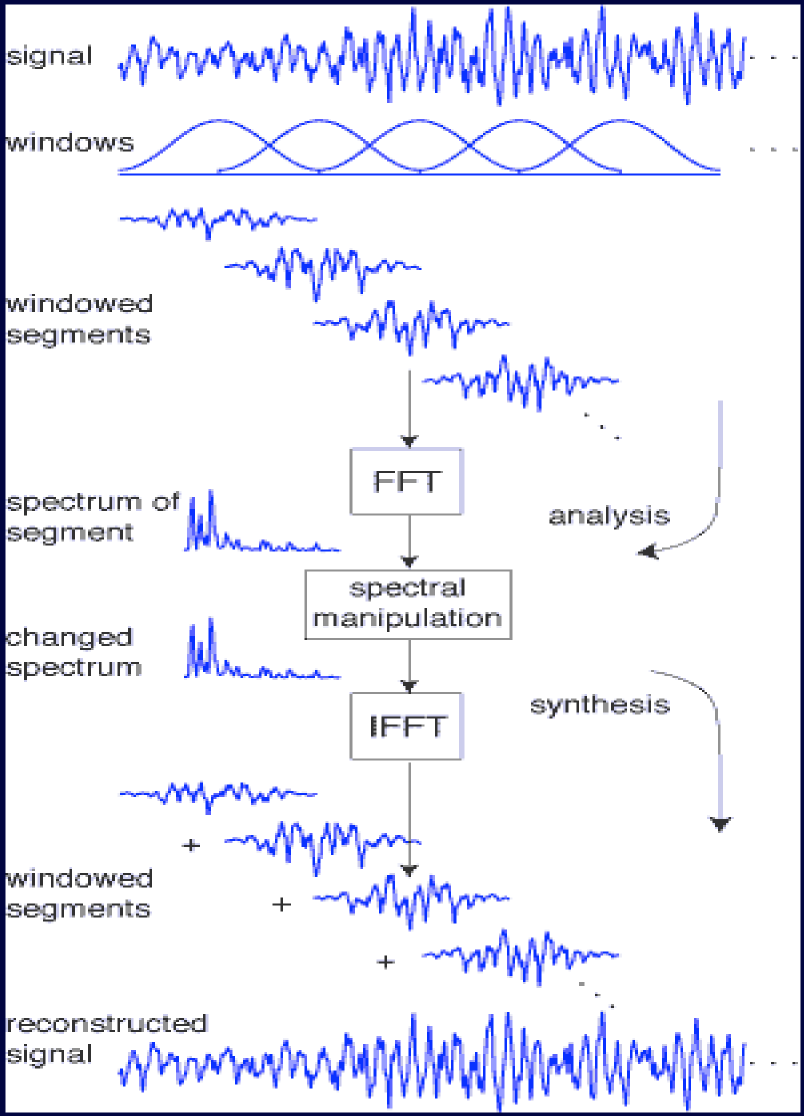
180, 225, 270, 315, 0, 45, 90
Need to convert to:
180, 225, 270, 315, 360, 405, 4

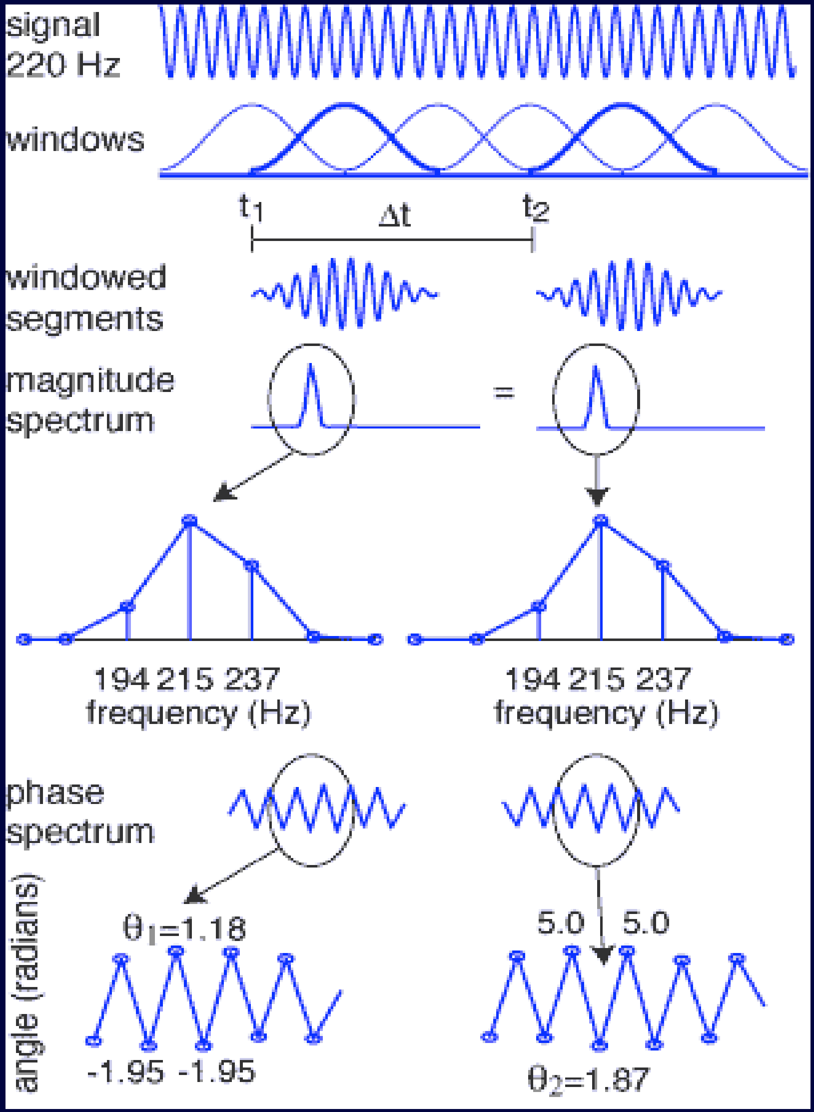
Phase unwrapping

Demos (waterfall display + phasevocoder)



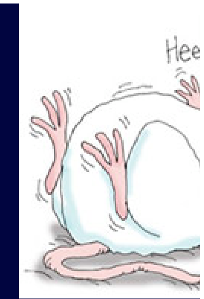
PhaseVocoder





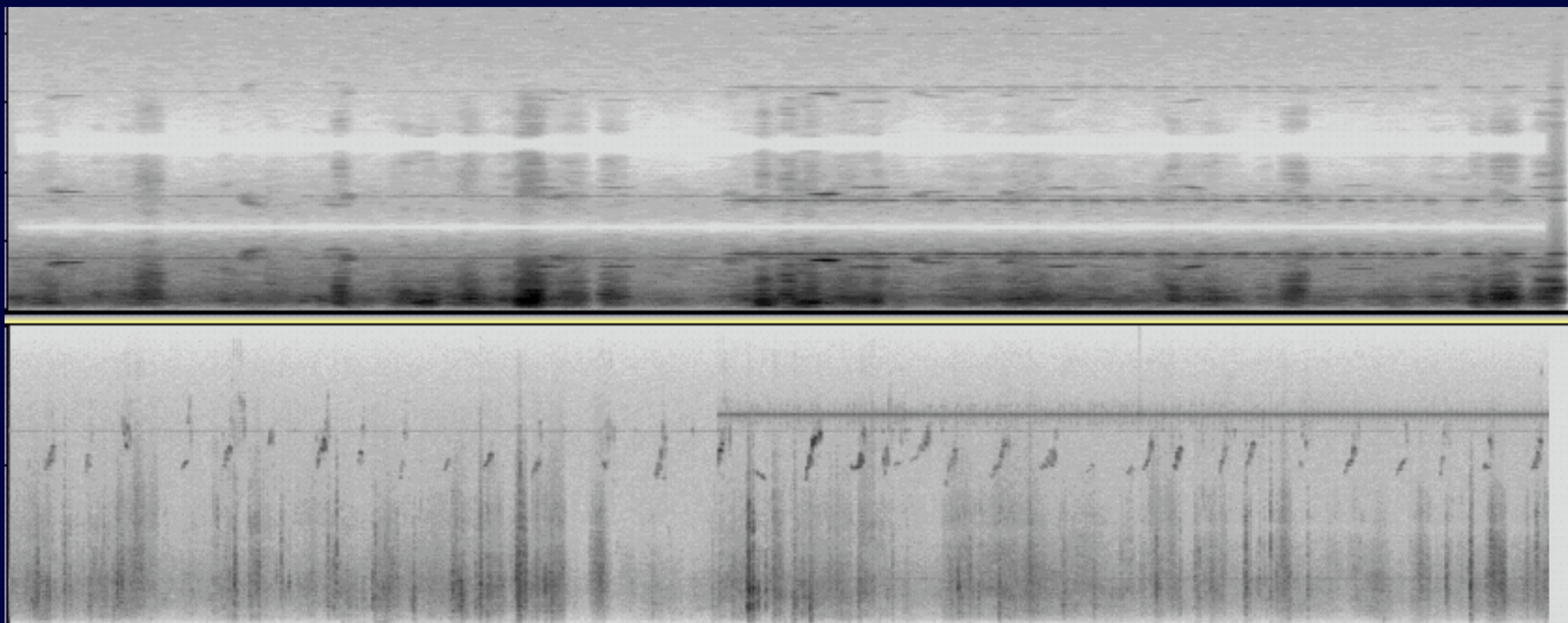
MarPhasevocoder Demo

Rats are ticklish



- Joint work with bio-acoustician Tecumseh Fitch (Univ. Viena, Austria)
- Rats laugh when tickled but we can't hear them - ultrasound range (50 KHz)
- Bat detectors simply frequency shift the signal destroying all harmonic structure
- Use phasevocoder for harmonic preserving pitch shifting and time-stretching (different metabolism)

Pitch Shifting

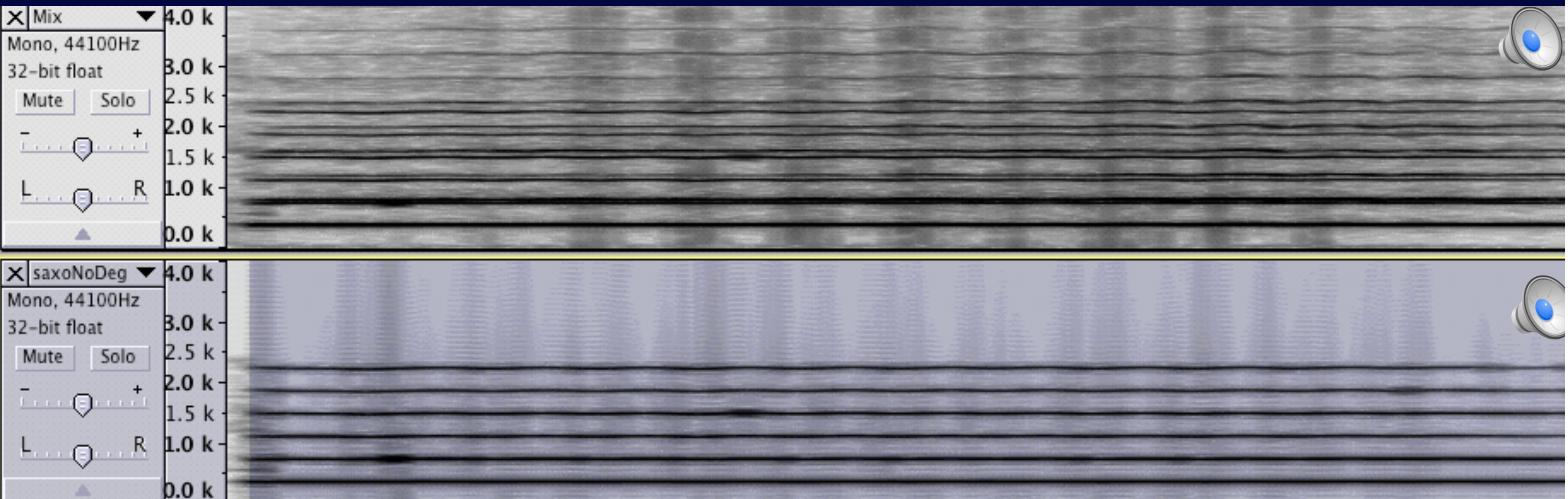


Original - what you hear is nail scratches

Pitch-shifted so that laughter is audible

Only laughter

Sound Source Separation



Gestalt Grouping Cues from Auditory Scene Analysis:

Frequency continuity

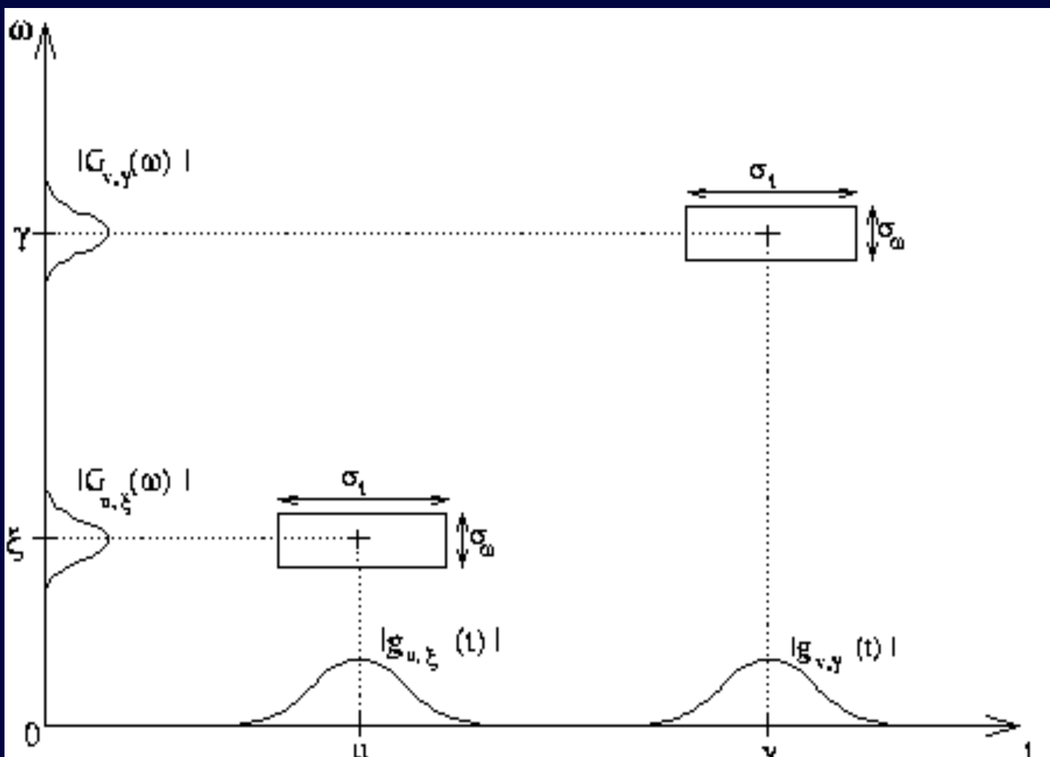
Amplitude continuity

Common Onset

Harmonicity

Time-Frequency tradeoff

FT = global representation of frequency content



Time – Frequency

L2 Heisenberg uncertainty

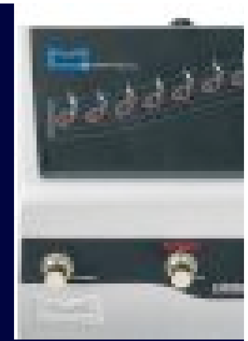
Time-Frequency

- Problem: to have more accurate frequency estimation we need larger windows of time
- Larger windows of time mean we don't know exactly when those frequencies happened
- What about making hopsize 1 ?
- Pitch-synchronous overlap-add (monophonic sounds)
- Holy grail = low latency polyphonic pitch shifting

IVL Audio Inc.

- Local company - products in professional music, consumer electronics, commercial karaoke
- Since 1984 - early pitch trackers)
- Late 80s pitch shifting
- Fall 2008 - DJ products

Morpheus Drop-tune



- Drop-tuning - retuning guitar strings to play lower (less string tension, multiple guitars)
- Morpheus drotune = guitar pedal for drop-tuning (marketed by XP audio, designed by IVL)
- Challenge = low latency polyphonic pitch shifting
- Based on phase-vocoding at multiple time resolutions
- Various youtube demos

"Classic" multi-stage approach

Time-Frequency representation

↓

Grouping Cue 1

↓

Grouping Cue 2

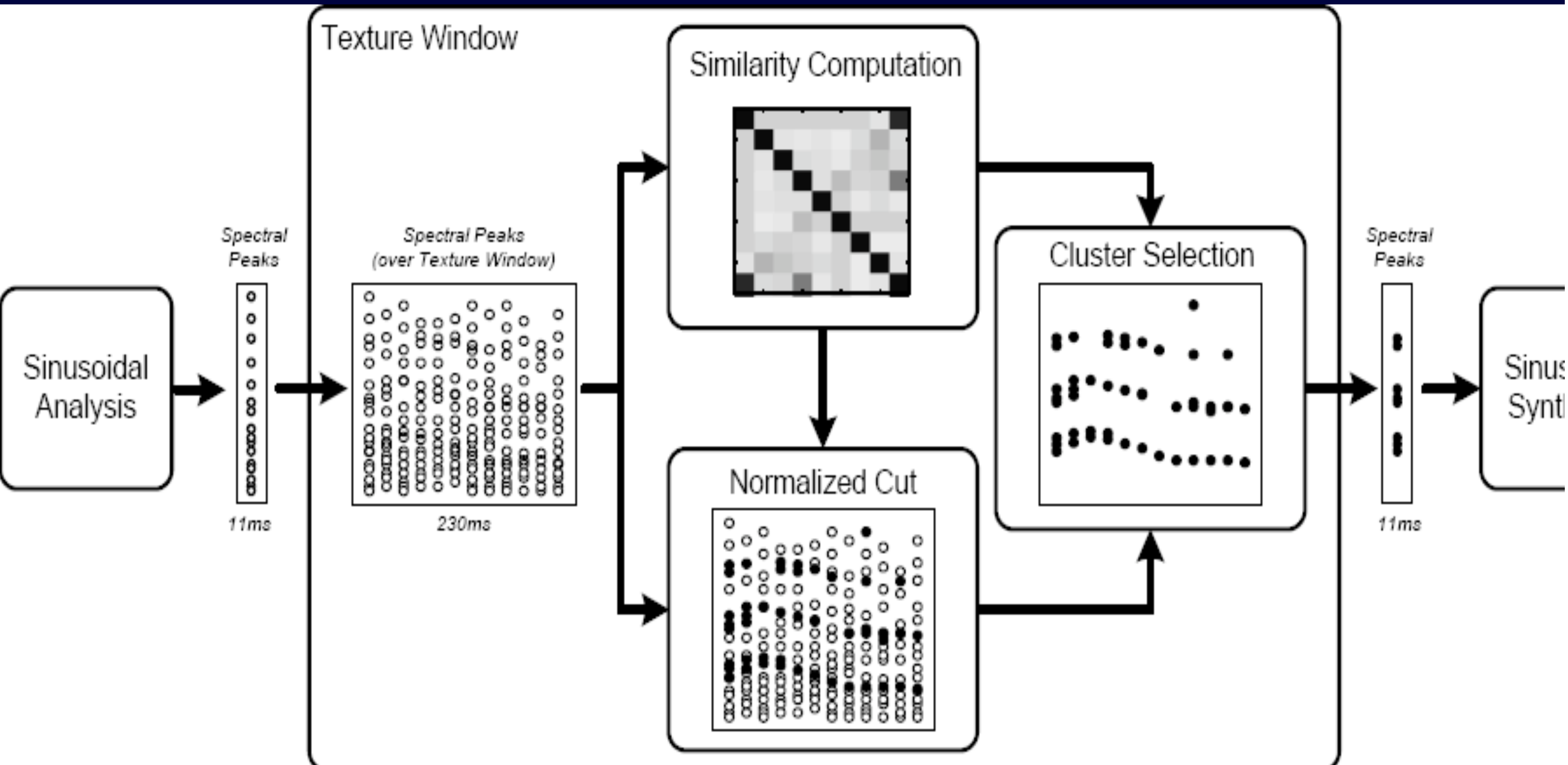
Short Time Fourier Transform
Discrete basis: windowed sine waves

Partial Tracking (McAuley & Quatieri)

Sound source formation:
grouping of partials based on harmonicity

PROBLEMS: Difficult to decide ordering, brit

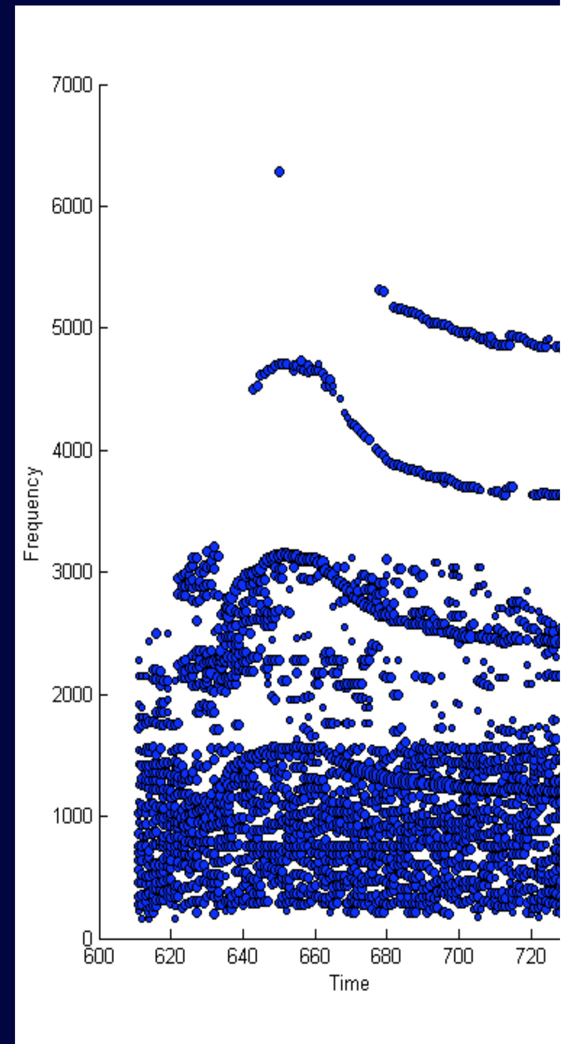
Sound Source Separation using Spectral Clustering



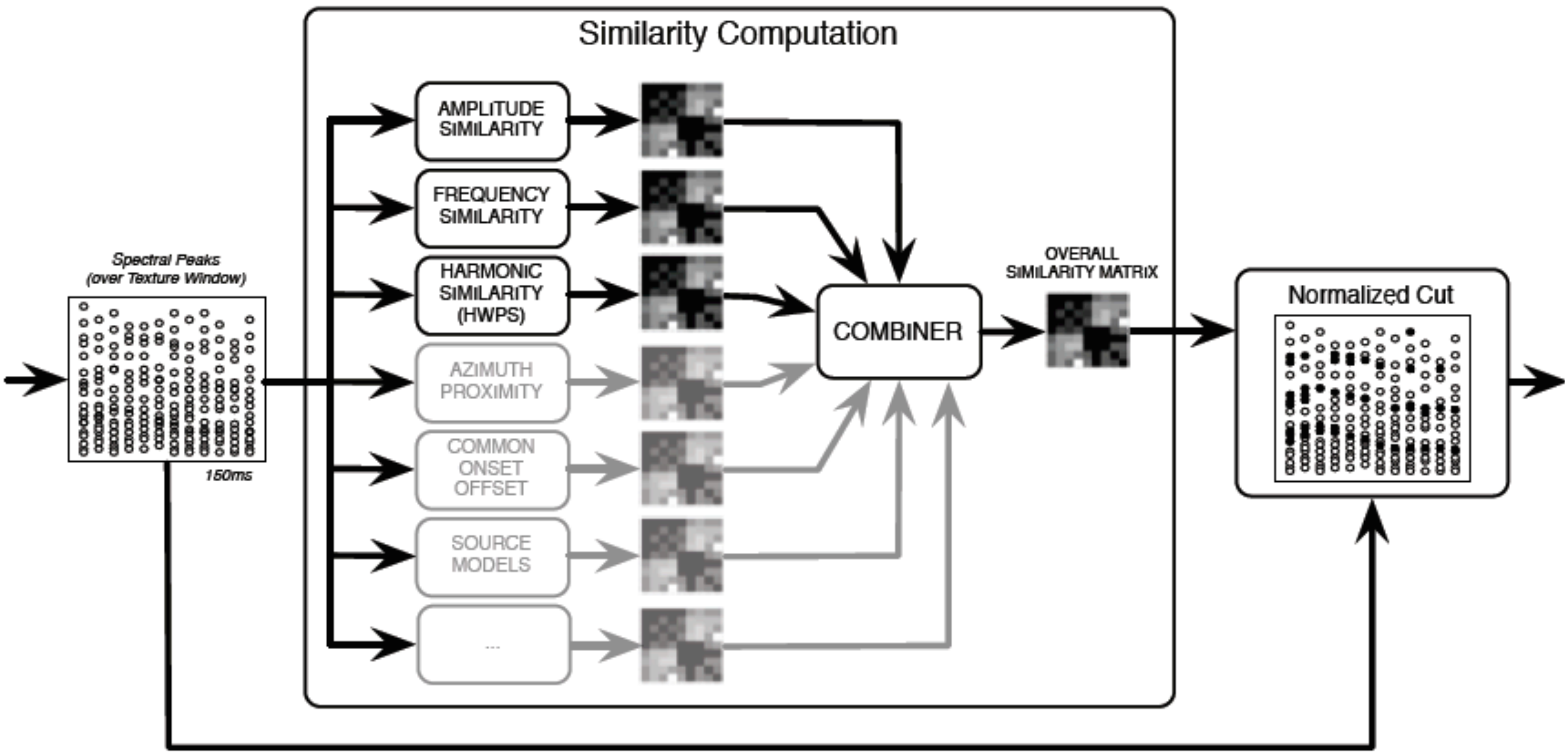
Sinusoidal Front-End

- The signal within a frame is modelled as:
  - peaks = sinusoidal components with constant parameters

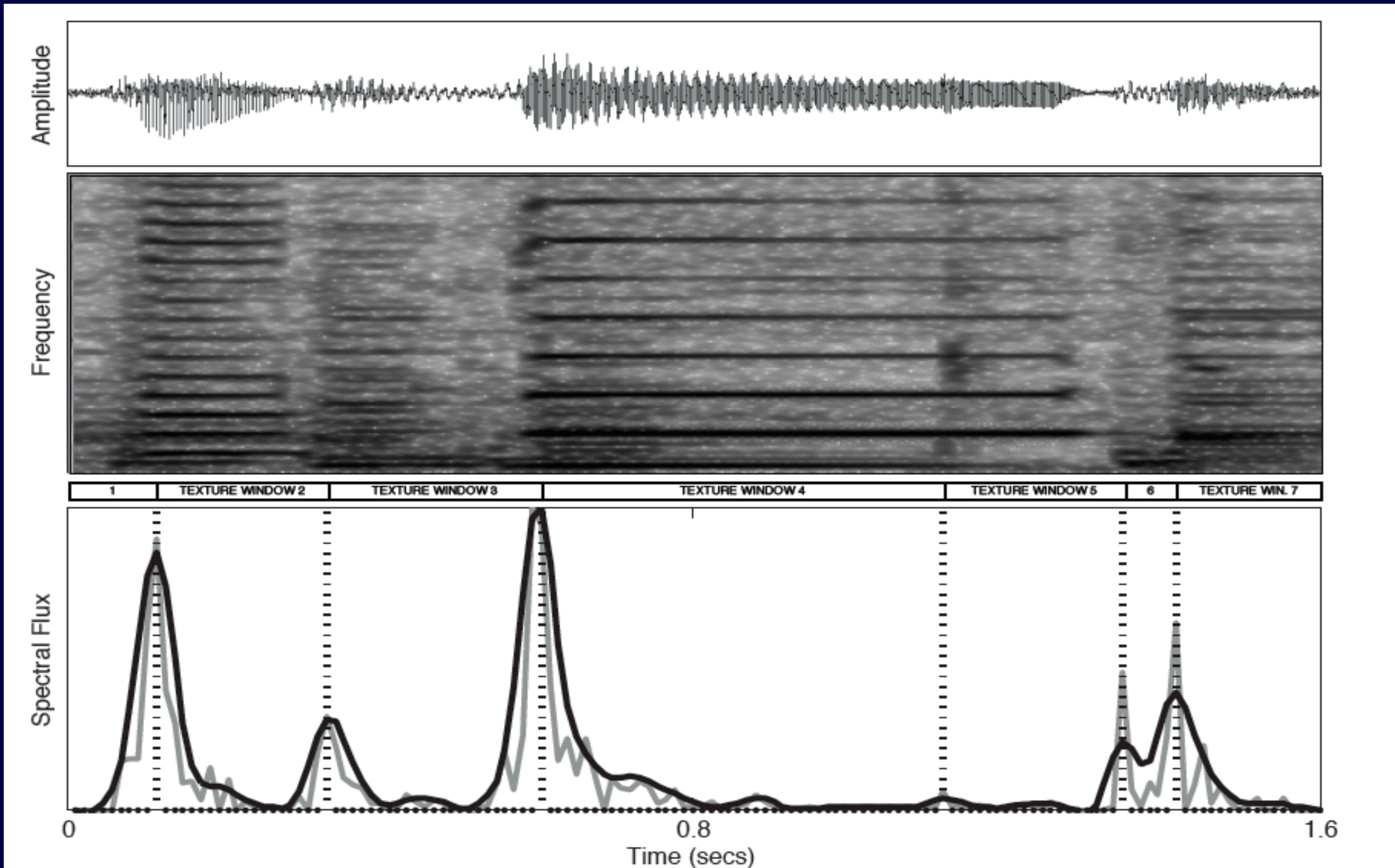
$$x^k(n) = \sum_{l=1}^{L^k} a_l^k \cos\left(\frac{2\pi}{F_s} f_l^k \cdot n + \phi_l^k\right)$$



Architecture

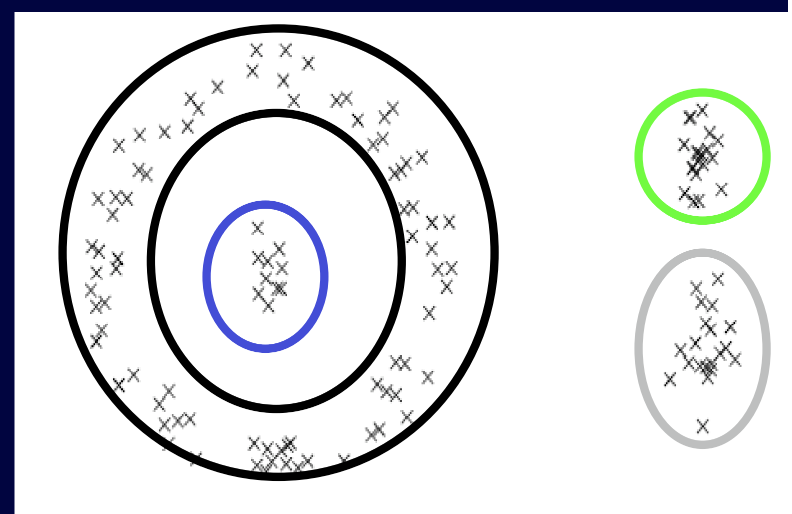


Dynamic texture windows

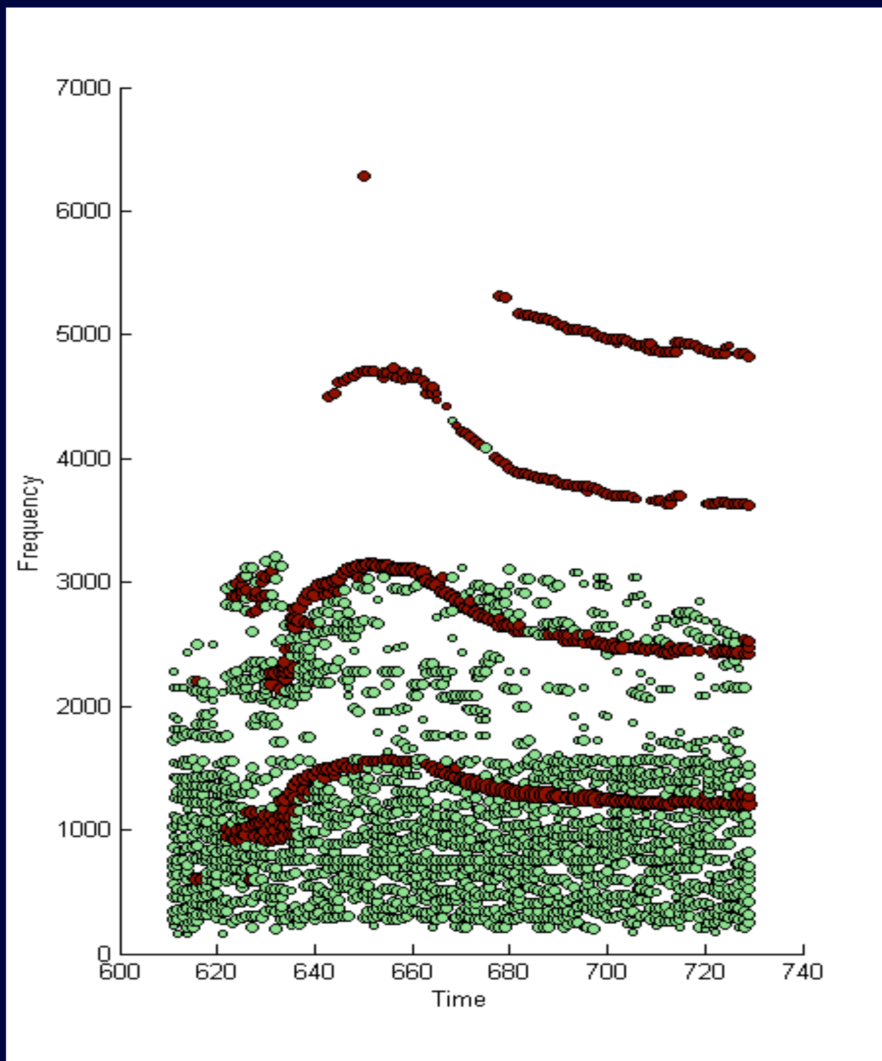


Spectral Clustering

- Alternative to traditional point-based algorithms (k-means)
- Doesn't assume convex shapes
- Doesn't assume Gaussians
- Avoid multiple restarts
- Eigenstructure of similarity matrix



Denoising of Orca Vocalizaton

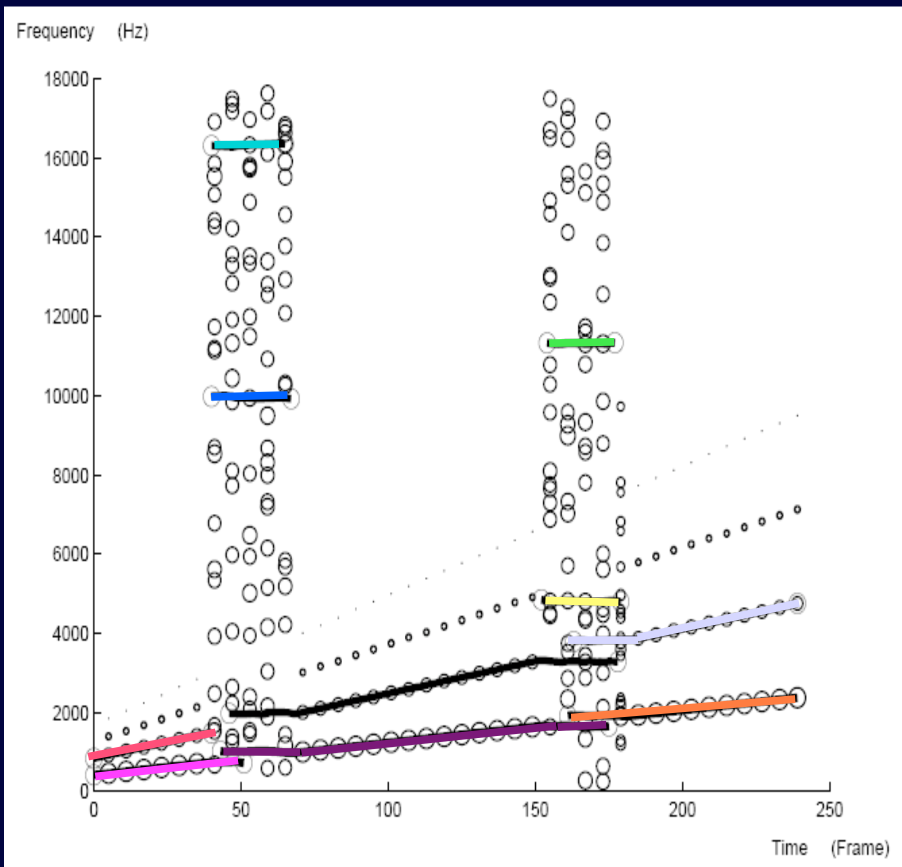


Original

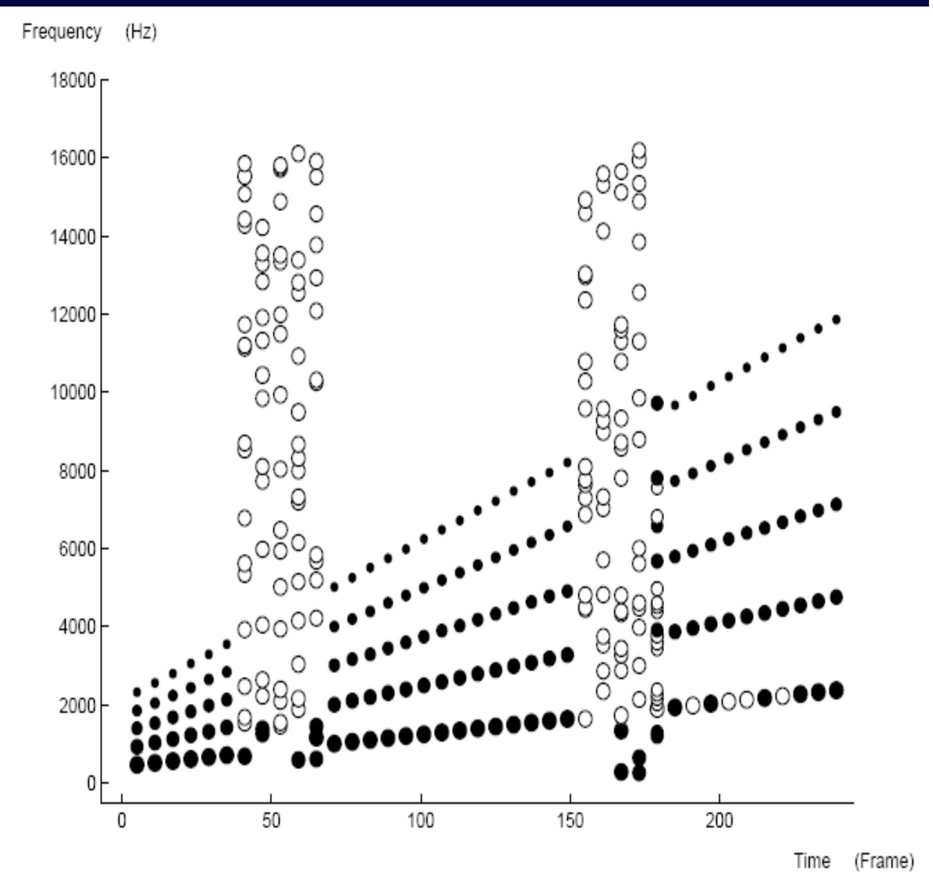
Noise

Separated Vocalization

Comparison with partial tracking

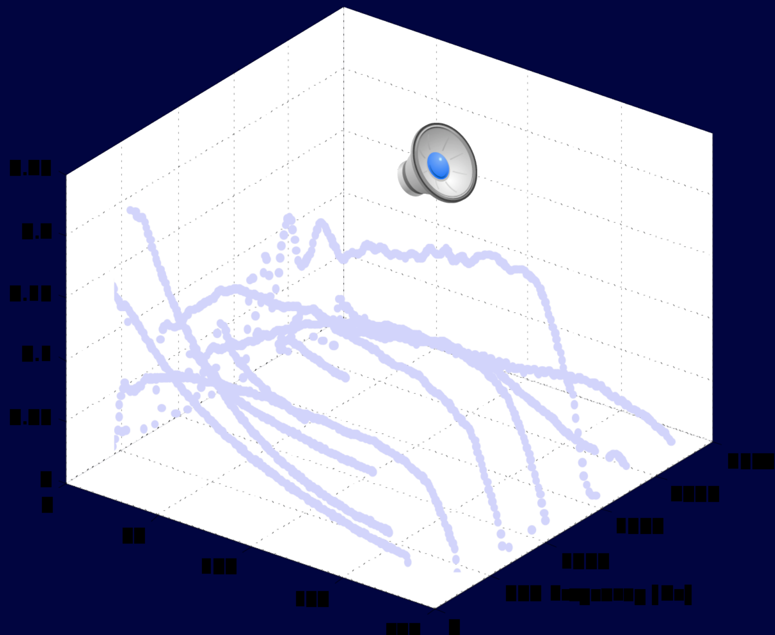


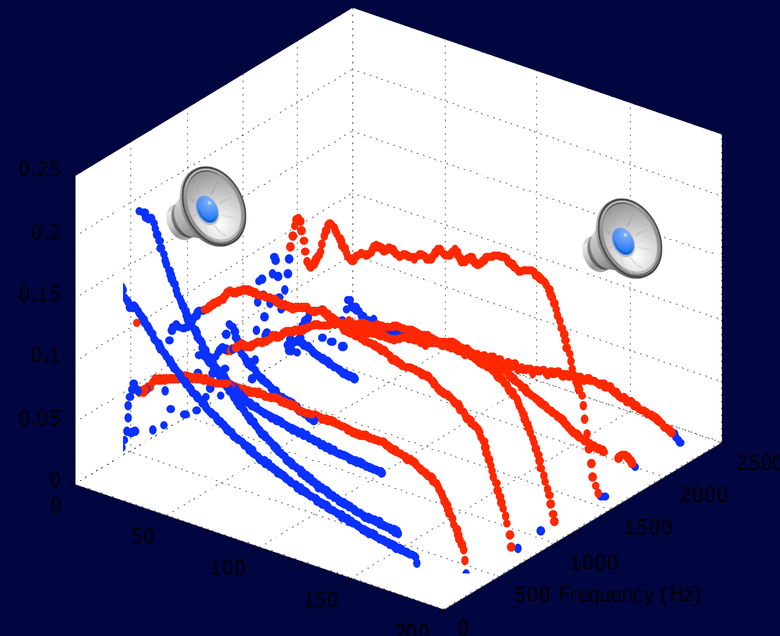
MacAuly and Quatieri
Tracking of Partials



Proposed Approach

Synthetic Mixtures of Instruments





4-note mixture

Instrumentation detection based on timbral models

Martins, et al, ISM

"Real world" separation results

Original Separated

- Mirex database
- Live U2

More examples: http://opihi.cs.uvic.ca/NormCutAudio
http://opihi.cs.uvic.ca/Dafx2007