CCRMA MIR Workshop 2012
Signal Analysis & Feature Extraction

Leigh M. Smith
iZotope Inc.
lsmith@izotope.com

Outline

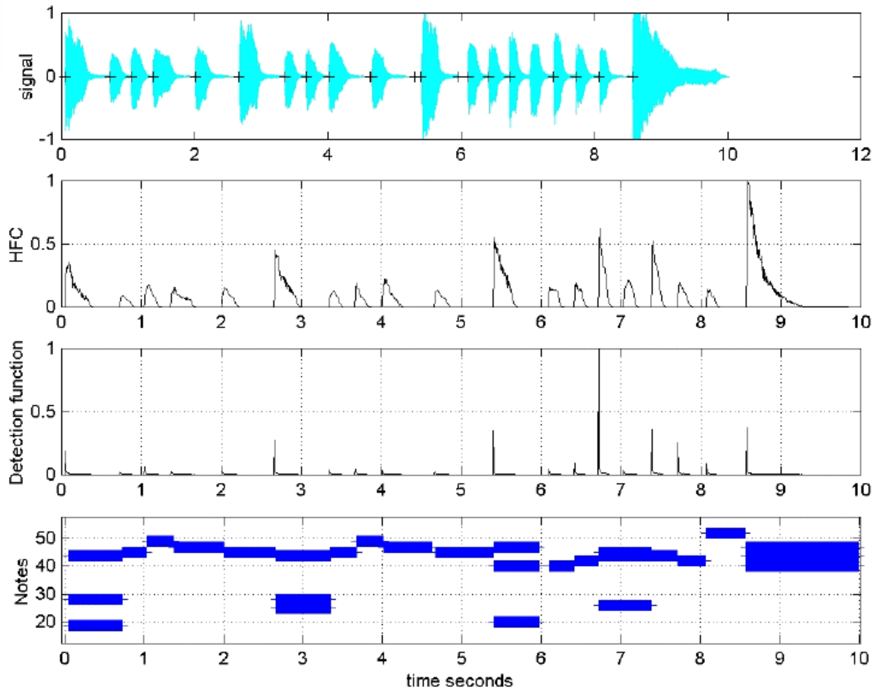
- MIR Applications
- Signal Analysis and Feature Extraction
- MIR Application Design
- Feature-vector Design
  - Time-domain Features
  - Windowed Feature Extraction
  - Frequency-domain Features
  - Spatial-domain Features
  - Other Feature Domains
- APIs

Introductions, Context

- Leigh Smith
  - Comp Sci Dept. University of Western Australia
  - Universiteit van Amsterdam – EmCAP project.
  - IRCAM – Quaero project.
  - Imagine Research Inc. now part of iZotope Inc.
  - CCRMA MIR Workshop 2011
- lsmith@izotope.com
- http://www.leighsmith.com/Research

Problem Statement: MIR Applications

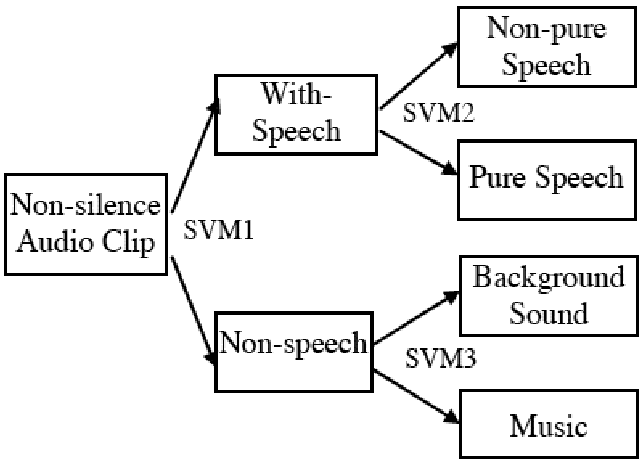
- Examples
  - Automatic playlist generation
  - Audio transcription



Courtesy Stephen T. Pope

Example App:

- Classification



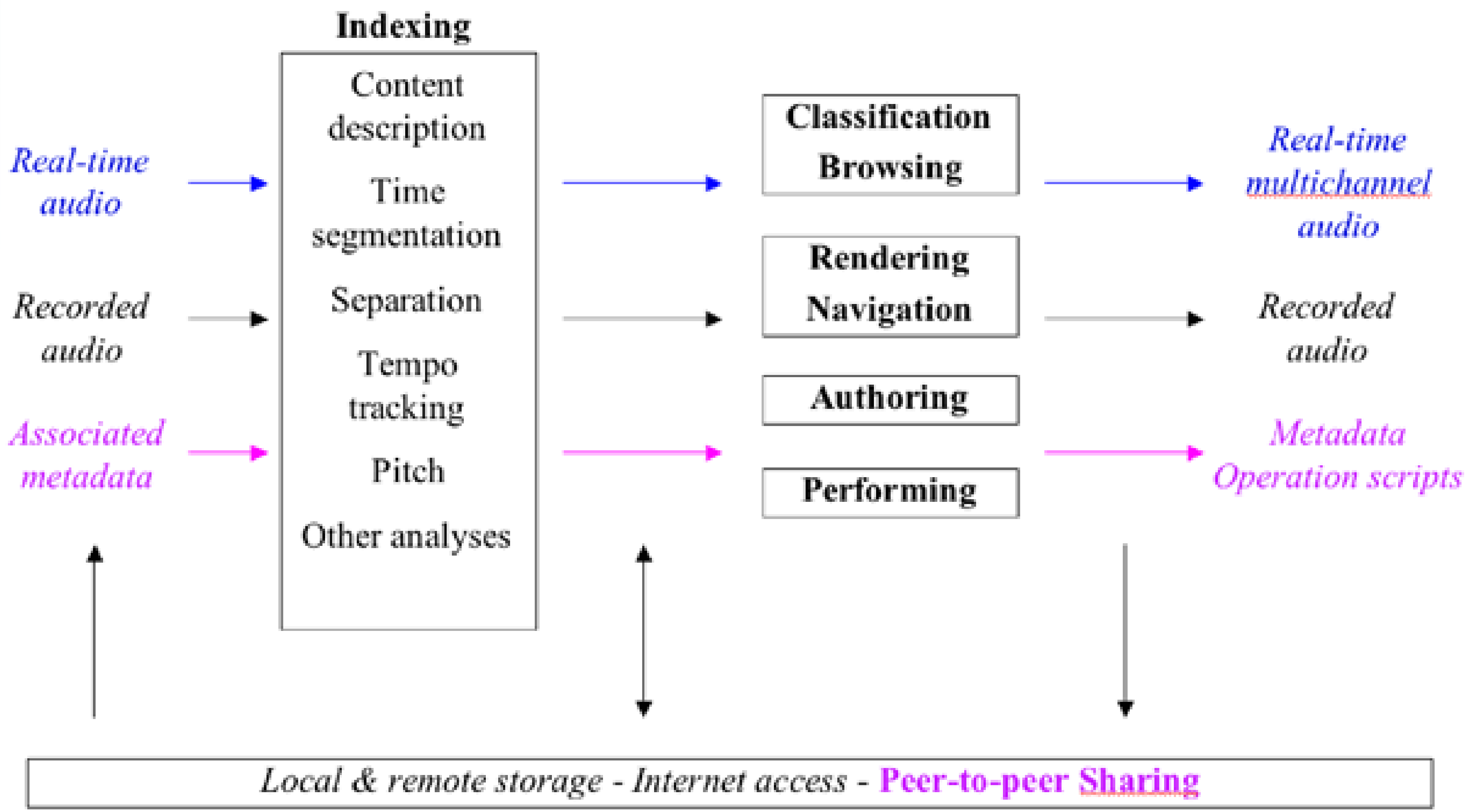
Signal Analysis and Feature Extraction for MIR Applications

- What do we want to do?
  - Match, search, index, transcribe, src-sep, ...
- What do we need to know to do it?
  - Basic feature set
  - Higher-level features
  - Feature data post-processing
  - Application integration
- MIR application design
  - How does the metadata fit in?
- Feature vector design for applications

Dimensions of Music Information Retrieval Applications

- Indexing, query, access
  - Use content or metadata for query
- Understanding, transcription
  - Derive (music/speech) model
- Clustering, classification
  - Feature vector for discrimination
- Content identification, finger-printing
- Preference-matching, recommendation

MIR/MDB Applications

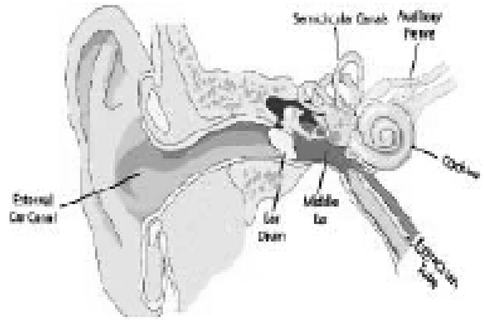


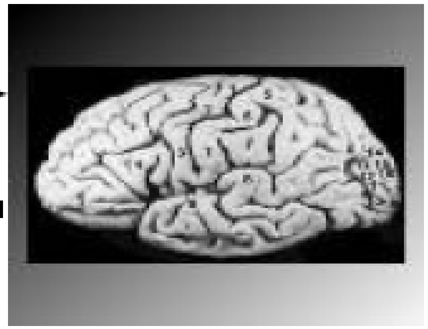
Tzanetakis's "MIR Pipeline"

Hearing
Representation

Understanding
Analysis

Acting
Interaction



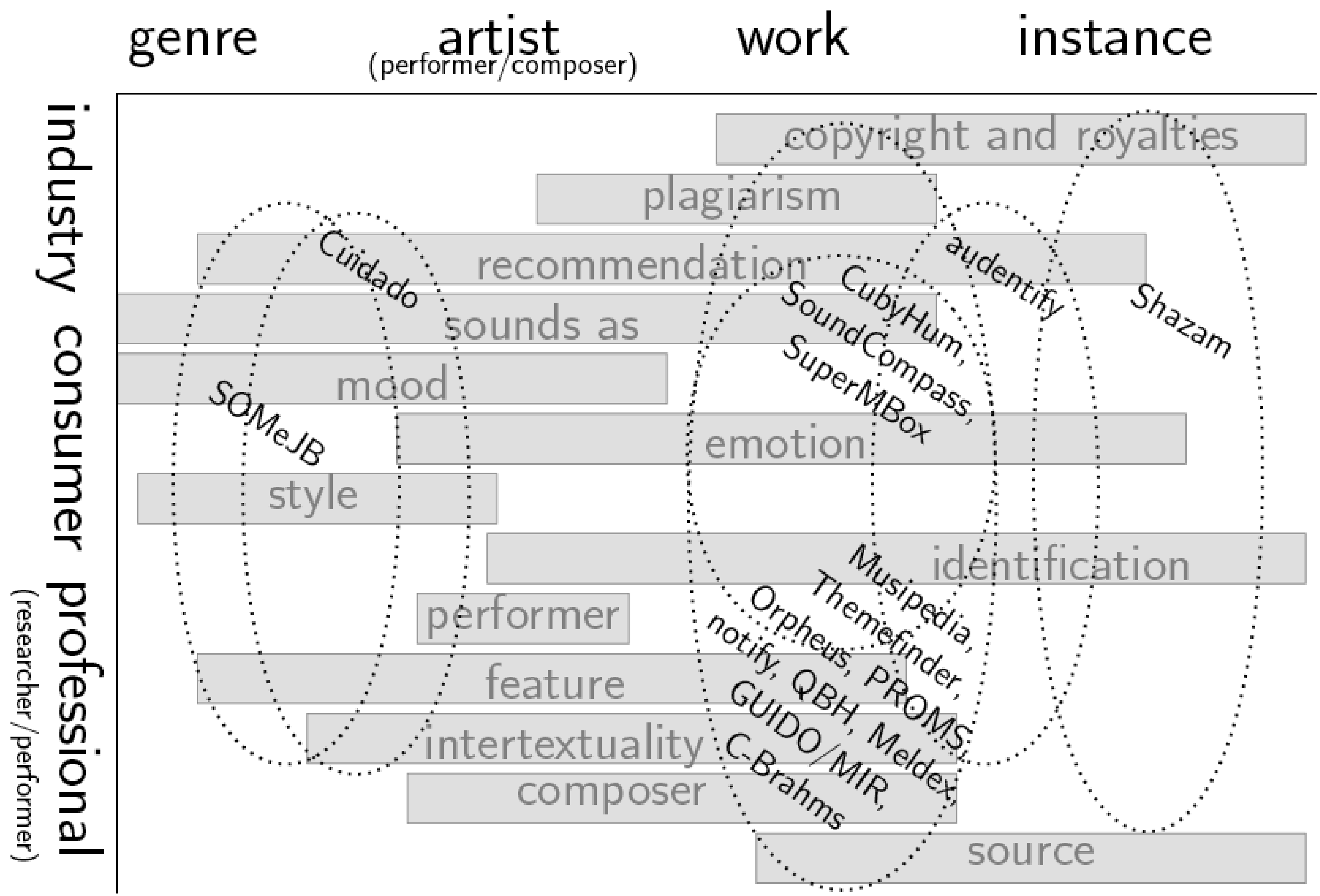


Signal Processing

Machine Learning

Human Computer
Interaction

MIR Systems & Retrieval Tasks

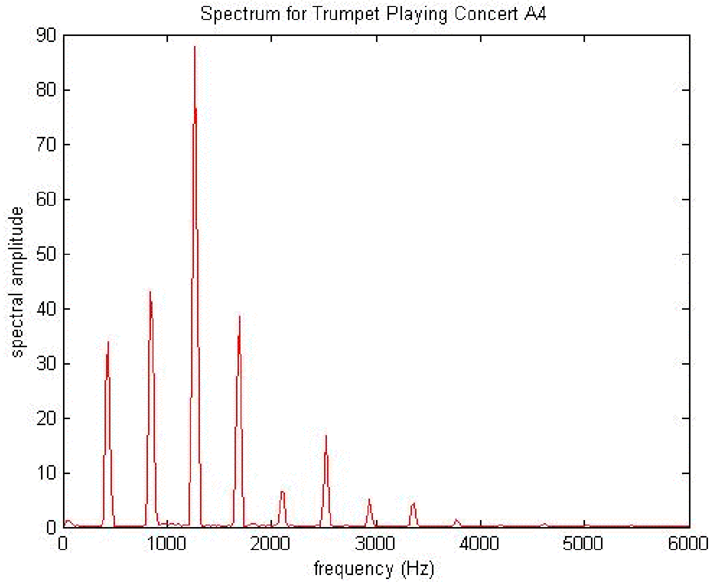


Monday, June 25, 2012

Typke's Appl. Table

Name	Input		Matching					Features								Indexing	Collection Size (Records)
	Audio	Symbolic	Audio	Symbolic	Exact	Approximate	Polyphonic	Audio Fingerprints	Pitch	Note Duration	Timbre	Rhythm	Contour	Intervals	Other		
audentify!	•		•			•	•	•								Inverted files	15,000
C-Brahms		•		•	•	•	•		•	•		•		•		none	278
CubyHum	•			•		•								•		LET	510
Cuidado	•		•			•	•				•	•			•	not de-scribed	works for $> 100,000$
GUIDO/ MIR		•		•		•			•	•		•		•	•	Tree of transition matrices	150
Meldex/ Greenstone	•	•		•		•							•	•		none	9,354
Musipedia	•	•		•		•							•			Vantage objects	$> 30,000$
notify! Whistle	•	•		•		•			•			•				Inverted files	2,000
Orpheus		•		•		•	•		•	•		•		•		Vantage objects	476,000
Probabilistic "Name That Song"		•		•		•								•	•	Clustering	100
PROMS		•		•	•	•			•			•				Inverted files	12,000
Cornell's "QBH"	•			•		•							•			none	183
Shazam	•		•		•		•	•								Fingerprints are indexed	> 2.5 million
SOMeJB	•		•			•	•								•	Tree	359
SoundCompass	•			•		•			•			•				Yes	11,132
Super MBox	•			•		•			•			•				Hierarchical Filtering	12,000
Themefinder		•		•	•				•				•	•		none	35,000

MIR Application Design Stages



- Considerations
  - Content format
  - Low-level analysis procedures
  - High-level derived features
  - DB design
  - Application flow and integration
- Design Issues
  - System architecture and design impacted by each of these decisions.

Content Format

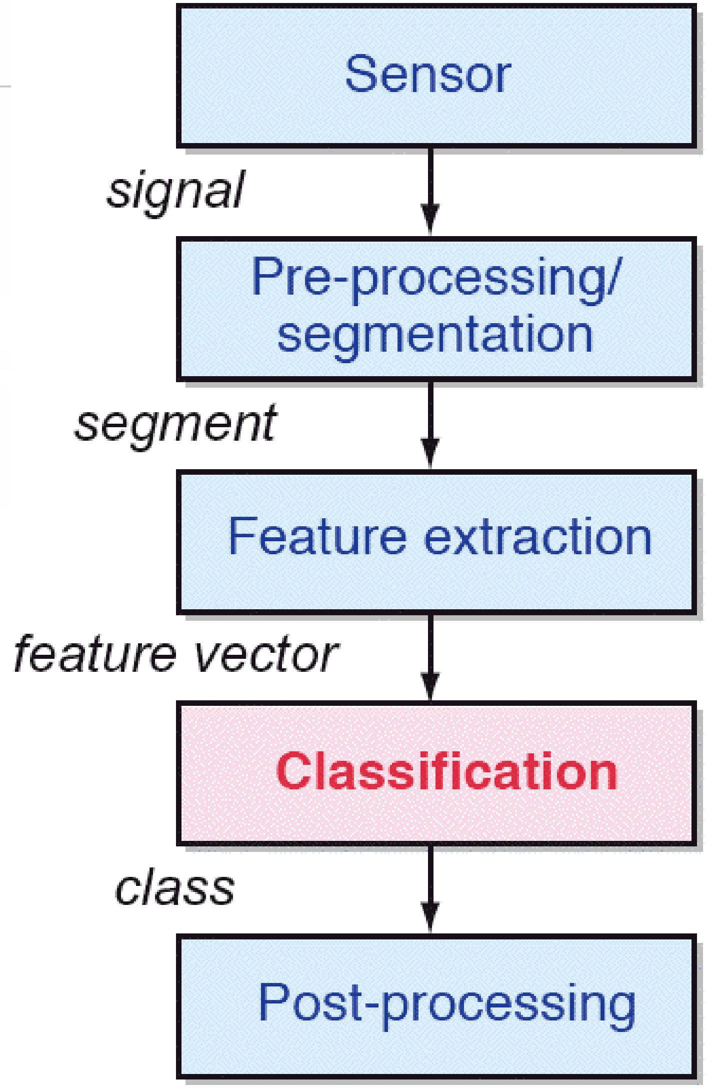
- Impacts all levels of system
  - Data volume, storage options, analysis DSP, DB design, etc.
- Systems may or may not maintain original source content (vs. metadata)
- Systems may preserve several formats of source and metadata (n-tier)
- This is typically a given rather than a design option

Content Formats

- Audio-based
  - Properties/volume of source recordings
  - MP3/AAC/WMA decoders
- MIDI-based
  - Problems with MIDI, assumptions to make
  - Human-performed vs “quantized” MIDI
- Score image based
  - Useful, but not treated here - genre specific.
- Formal language-based
  - SCORE, SMDL, Smoke, etc.
  - MusicXML

Numerical Processing

- Data Reduction, Smoothing
- Correlation, Grouping
- Princ./Indep. Component Analysis
- Audio Segmentation and Musical Form
- Clustering and Classification

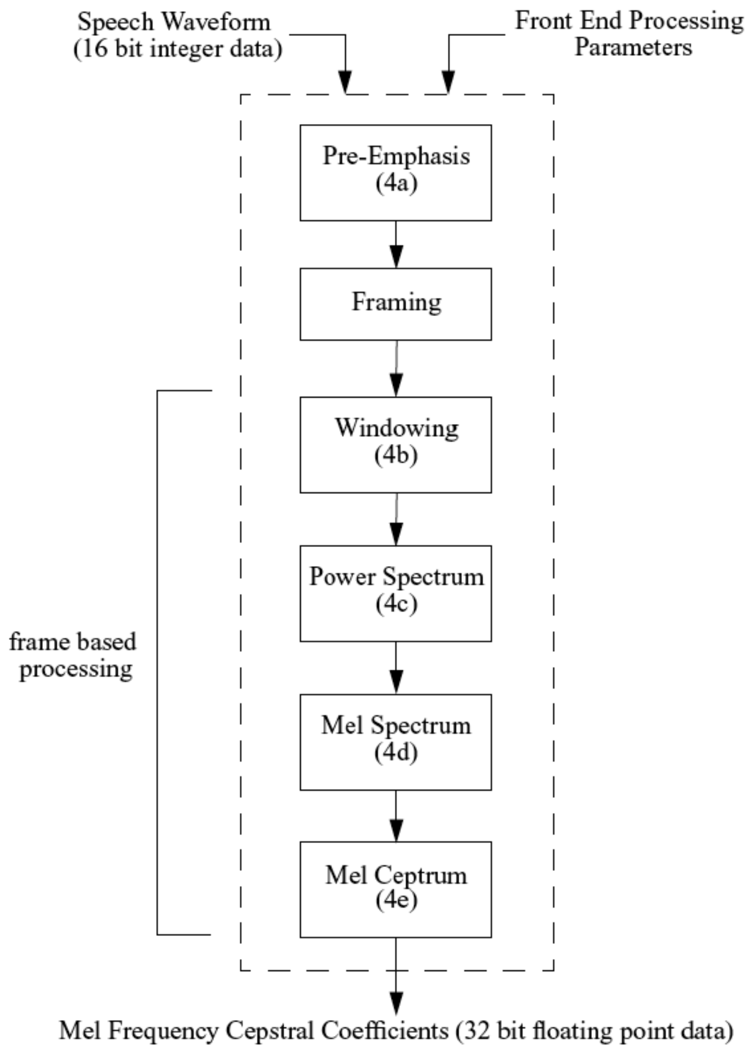


Typical Processing Stages

- Input processing
  - Streaming, decompression, reformatting
- Signal segmentation, windowing in time/freq
  - window size, share, overlap
- 1st-pass windowed feature extraction
  - Basic time-, freq-domain features
- 2nd-pass feature processing
  - Feature massaging, smoothing, pruning
  - 2nd-pass features (tempo, segmentation)
- Post-processing, data output
  - Many options

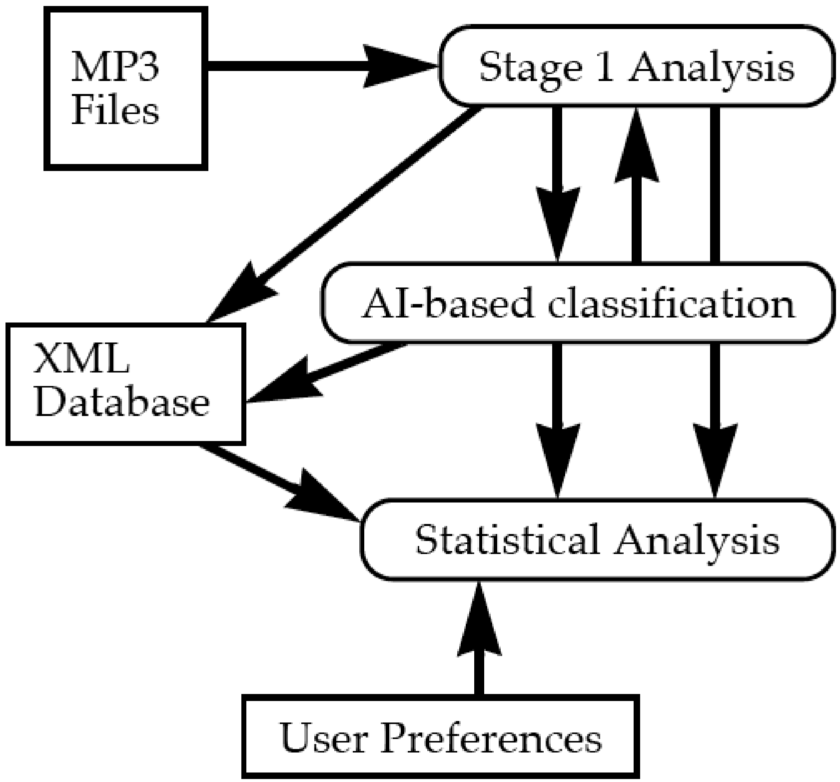
Signal Analysis

- Time-domain Audio Analysis
  - Windowed RMS Envelope Extraction
  - Beat Detection and Rhythm Analysis
  - Time-based signal segmentation
- Frequency-domain Analysis
  - Pitch Detection Techniques
  - Spectral Analysis and Interpretation
  - Spectral Peaks and Tracking
  - Other Spectral Measures
- Other Kinds of Analysis: Wavelets
- Cross-domain analysis

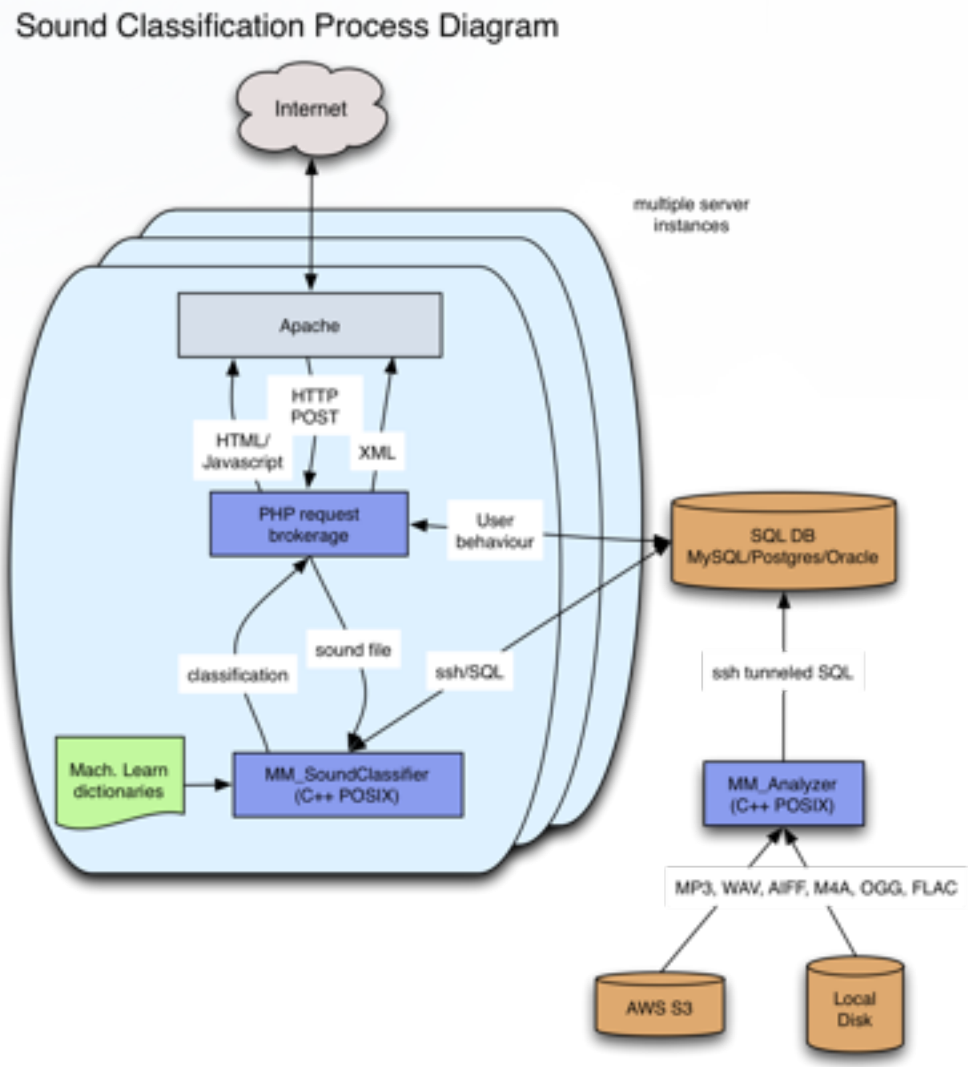


Real Applications

- Query systems, browsers, and MIR frameworks
- DBMS issues
- Machine Learning
- Informed tools
- Stand-alone delivery applications



Implementation example



Databases & Applications

- Searching, Indexing, and Players
- Audio Summarization and Thumb-nailing
- Content Matching and Finger-printing
- Data Clustering and Genre Classification
- Other Applications

Database Technology

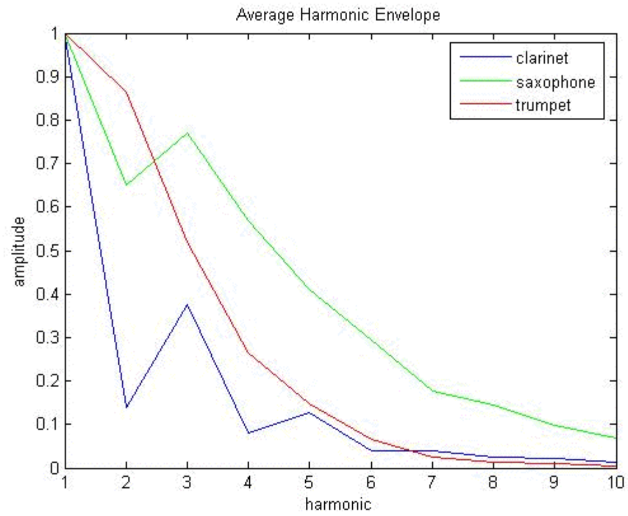
- Database Designs: Schema vs. None?
- Relational DBMS (MySQL/Oracle/PostgreSQL)
  - Fixed table-formatted data
  - Few data types (number, string, date, ...)
  - One or more indices/table (part of DB design, application-specific, impacts performance)
  - Cross-table indexing and joins
  - SQL examples (create, insert, update, select)
- Media data (historically images)
  - Volume (large single items)
  - Format (items no known structure)
  - Content and metadata (required for usage)
  - Handling of Large/Dynamic Feature Vectors (MongoDB)
- Consider Application Requirements and Design

Feature-vector Design

- Application Requirements
  - Labeling, segmentation, etc.
  - Derive feature vector from the app requirements
- Kinds/Domains of Features
  - Time-domain
    - Simple features, onset detection
    - Rhythm, segmentation
  - Frequency-domain
    - Spectrum, spectral statistics
    - Pitch, chroma, key

(See e.g: http://www.create.ucsb.edu/~stp/PostScript/PopeHolmKouznetsov_icmc2.pdf)

Feature Vectors and Indexing



Feature = derived (numerical) parameter

- **Feature vector** = list of features for a single point/window in time, or average for an entire selection
- **Feature table** = list of feature vectors for several time slices (not always used/stored)

Example Features

- Features:
  - Time-domain, low-level
    - Windowed RMS amplitude
  - Time-domain, high-level
    - Tempo, beat structure, segmentation
  - Frequency-domain, low-level
    - Pitch, spectrum, spectral peaks
  - Frequency-domain, high-level
    - Peak track birth/death statistics, instrument ID
  - Many other possibilities (see below)

Feature Vector Examples

Product Comparison

Field	Bringin' Da Noise	I'll Be Your Everyth...	Weighted
Volume Width	48.126621	47.903584	0.182064596871...
LPC Avg-Track-Dur	260.071	291.654	0.246736659056...
Bass Loudness	-3.82097	-3.48169	1.151592910141...
Spectral Contrast	17.8124	27.7138	1.260984294687...
LPC Track-Harmo...	1.15606	1.10925	1.386355020613...
BusyMid	399.87873138	382.9394489400...	2.090529929650...
Freq Max	0.579932	0.629061	2.756166578401...
Average Volume	34.344021	37.742193	3.092778888824...
Freq Avg	0.004416	0.004209	3.273244781783...
Tempo	111.966	105.943	3.872166433080...
LPC Peaks-Per-S...	258.61	229.837	5.144795608229...
LPC Freq-Deviation	6257.06	5584.61	5.146495852036...
% Freq Over Avg	24.050509	21.898819	5.313072728419...
Spectral Variety	57.0208	97.2588	5.591531132924...
BusyLow	412.44579522	341.0040312499...	6.891936456624...
Spectral Saturation	0.712956	0.651703	7.476978442821...
LPC Tracks-Per-S...	56.5431	48.2628	7.601499754612...
Snare Strength	0.328855	0.235586	8.982285629537...
Overall Grunge	0.248330529671...	0.067614786427...	12.20650524954...
% Rhythm	99.48301435406...	97.82279545454...	N/A
BEAT: hiquot	5.2	5.8	N/A
BEAT: maxscore	1550.0	926.0	N/A
BEAT: spikewon	0.0	0.0	N/A
BEAT: window	20.0	20.0	N/A

Shapes Close

Example: FMAK3 Feature Table

```
class FeatureTable {                      // FeatureTable is a root object (no parents)

public:                                   // Data members (instance variables)
          float mTimeStamp;               // When do I start?
          float mTimeDur;                 // How long a time-span do I represent?
                                          // Time-domain features
          unsigned int mRMSWindowSize;    // Size of RMS window
          FeatureDatum mRMS;              // Rectangular-windowed RMS amplitude
          FeatureDatum mPeak;             // Max sample amplitude
          FeatureDatum mLPRMS;            // RMS amplitude of LP-filtered signal
          FeatureDatum mHPRMS;            // RMS amplitude of HP-filtered signal
          size_t mZeroCrossings;          // Count of zero crossings
          FeatureDatum mDynamicRange;     // RMS dynamic range of sub-windows
          FeatureDatum mPeakIndex;        // RMS peak sub-window index
          FeatureDatum mTempo;            // RMS/FWT instantaneous tempo estimate
          FeatureDatum mTimeSignature;    // Time signature guess
          FeatureDatum mBassPitch;        // Bass pitch guess in Hz
          unsigned int mBassNote;         // Bass note (MIDI key number) guess
          FeatureDatum mBassDynamicity;   // Bass note dynamicity (size of histogram)
                                          // Spatial features
          FeatureDatum mStereoWidth;      // L/R difference
          FeatureDatum mSurroundDepth;    // Front/Surround difference
          FeatureDatum mCenterDistinction; // Center vs. L/R sum difference
```

Example: FMAK3 Feature Table, cont'd

```
                              // Frequency-domain features
unsigned int mFFTWindowSize;  // Size of FFT window
FtVector mSpectrum;           // Hanning windowed FFT data (1024 points, or NULL)
FtVector mReducedSpectrum;    // 1-octave FFT data (10-12 points)
FtVector mBandSpectrum;       // 2.5-octave FFT data (4 points -- spectral bands)
FPartialVector mSpectralPeaks;// List of major spectral peak indeces
FPartialVector mSpectralTracks; // List of tracked peak frequencies
FeatureDatum mSpectralCentroid; // Spectral centroid measure
FeatureDatum mSpectralSlope;  // Spectral slope measure
FeatureDatum mSpectralVariety;// Inter-frame spectral variety measure
                              // Hi-frequency properties
FeatureDatum HiFreqBalance;   // Relative HF level
FeatureDatum HiFreqVariety;   // HF inter-frame spectral variety
FeatureDatum HiFreqCorrelation;// Correlation between HF and audio-band tracks
FeatureDatum mSTrackBirths;   // Spectral peak track births and deaths
                              // LPC features
unsigned int mLPCWindowSize;  // Size of LPC window
FPartialVector mLPCFormants;  // List of LPC formant peaks
FPartialVector mLPCTracks;    // List of tracked LPC formants
FeatureDatum mLPCResidual;    // LPC residual level (noisiness)
FeatureDatum mLPCPitch;       // Pitch estimate
FeatureDatum mLTrackBirths;   // LPC formant peak track births, deaths
                              // Wavelet-domain (FWT) features
FtVector mWaveletCoeff;       // FWT coefficient or NULL
FtVector mWTNSpectrum;        // Reduced FWT HiFreq noise spectrum
FtVector mWTTracks;           // List of tracked FWT peaks
FeatureDatum mWTNoise;        // FWT noise estimate
```

Intermission

- To be continued!

Analysis Domains and Transformations

- Time-domain Audio Analysis and Applications
  - Windowed RMS Envelope Extraction
  - Beat Detection and Tempo Analysis
  - Time-based signal segmentation
- Frequency-domain Analysis
  - Pitch Detection Techniques
  - Spectral Analysis and Interpretation
  - Spectral Peaks and Tracking
  - Other Spectral Measures
- Other Kinds of Analysis: Wavelets
- Cross-domain analysis

Feature Extraction and Signal Analysis

- Multi-step process:
  - Read input
  - Apply window
  - Derive several low-level features
  - Map, derive next-level features
    - Possible heuristics determine which next-level features are relevant
  - Prune data when appropriate

Time-domain Features

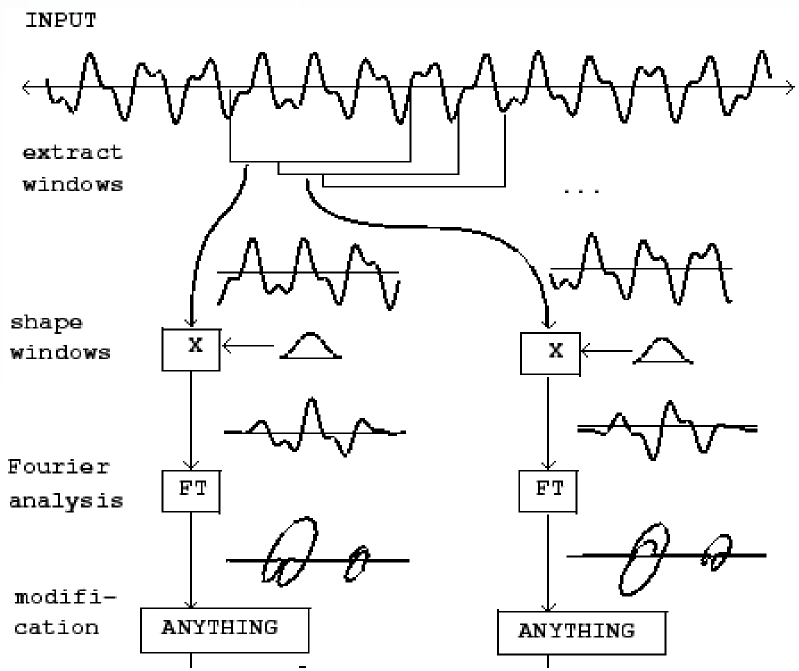
- RMS, Peak

$$RMS = \sqrt{\frac{1}{N}\sum_{n}^{N} x_n^2}$$

- LPF/HPF RMS
  - e.g., F < 200 Hz, F > 2000 Hz
- Dynamic range
  - What window for calc?
- Zero-crossing rate (time- or freq-domain?)
- Higher-level statistics
  - Mean/variance
  - Variance of sliding windows
  - Spacing of peaks/troughs
  - Many other options
- Time-domain onset-detection & beats

Time Sequences, Windowing

- Read audio input
- Vector multiply by window function
- Perform analysis
- Step to next window
- Hop size not normally = window size (overlap)
- Window features
  - Main lobe width, side lobe level, side lobe slope



Time-domain Audio Analysis and Applications

- Use rectangular window if no overlap or triangular window if overlapping
- Medium-sized window (10 Hz or better resolution desired)
- Derived windowed RMS value
- Count zero crossings

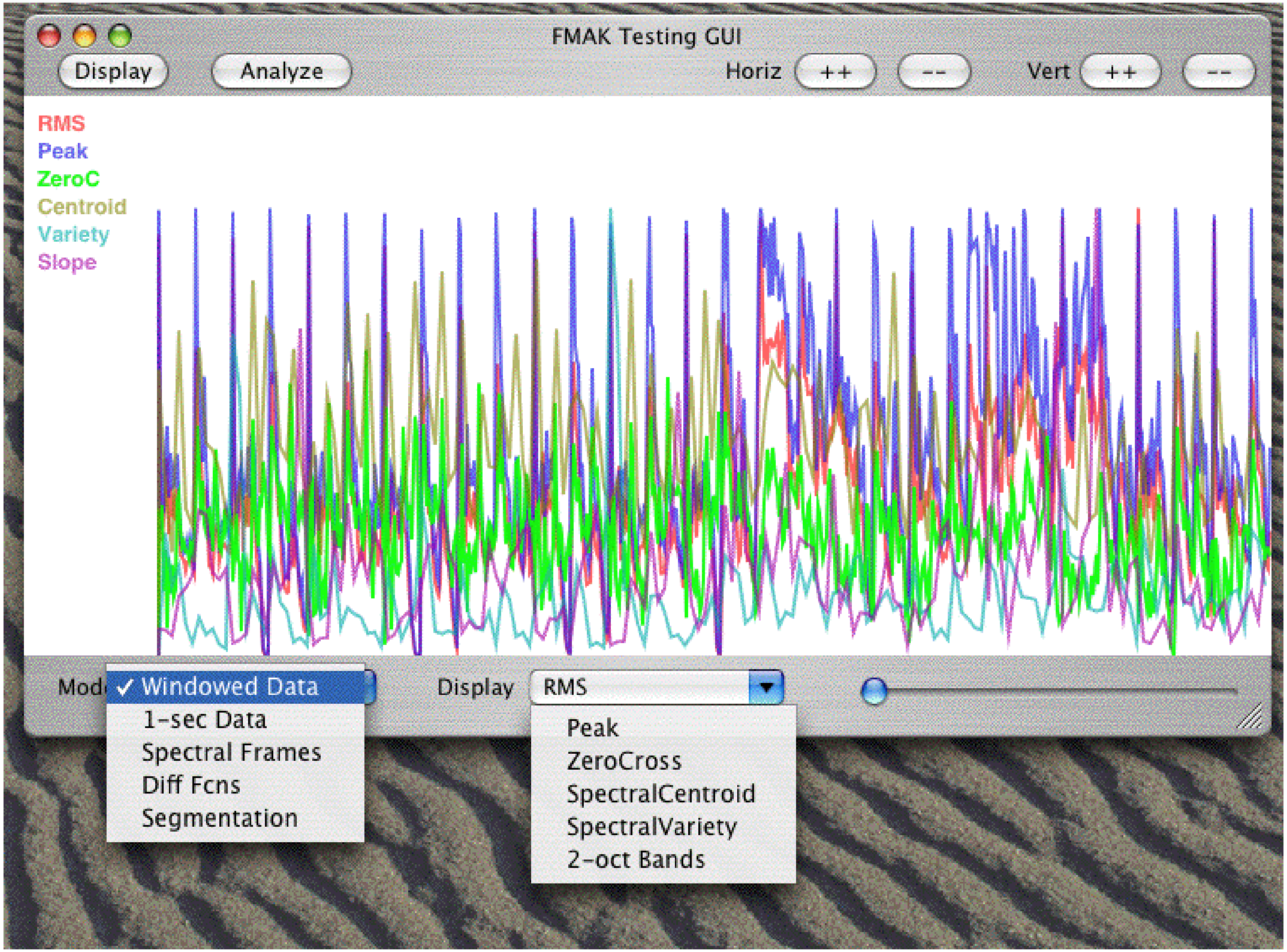
Windowed RMS Envelope Extraction

- pseudo-code for envelope extraction:
  - Outer loop for windows
  - Inner loop to run window and compute RMS value
  - Silence threshold (noise gate)
  - Note-on trigger (peak detector)
  - Example sound: piano sample, drum loop

Optional Time-domain Steps

- Pre-filter to get low-freq and high-freq RMS values
- Process stereo channels to get M/S (sum/difference) signals
- Noise detection
- Silence detection

Windowed Feature Comparison



Windowed Amplitude Envelopes

- Choice of window size, hop size, window function shape
- May use several frequency bands (kick drum vs. hi-hat)
- Useful for silence detection, beat tracking, simple segmentation, summarization, etc.
- Simple, effective, well-understood techniques, many options

Frequency-domain Features

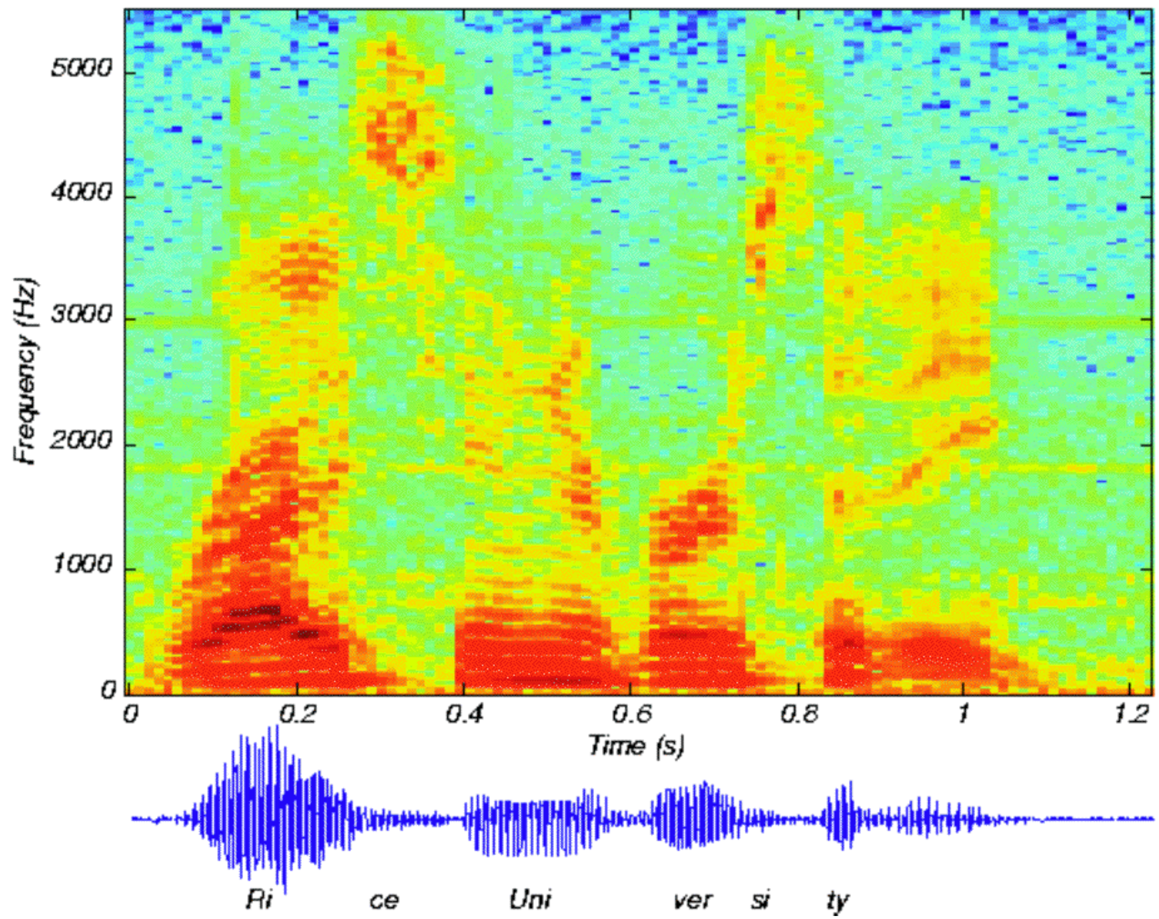
- Spectrum, Spectral bins
  - Window/hop sizes
  - Improving spectral data: phase unwrapping, time realignment
- Spectral measures (statistical moments)
- MFCCs
- Peak-picking and peak-tracking
- Pitch-estimation and pitch-tracking

Frequency-domain Analysis

- Short-time Fourier transform
  - Configuration options and trade-offs
  - Interpretation/weighting of spectral bins (perceptual scales)
- Other frequency-domain techniques
  - Filter banks
  - Linear prediction
  - Filter matching
- Loads of options

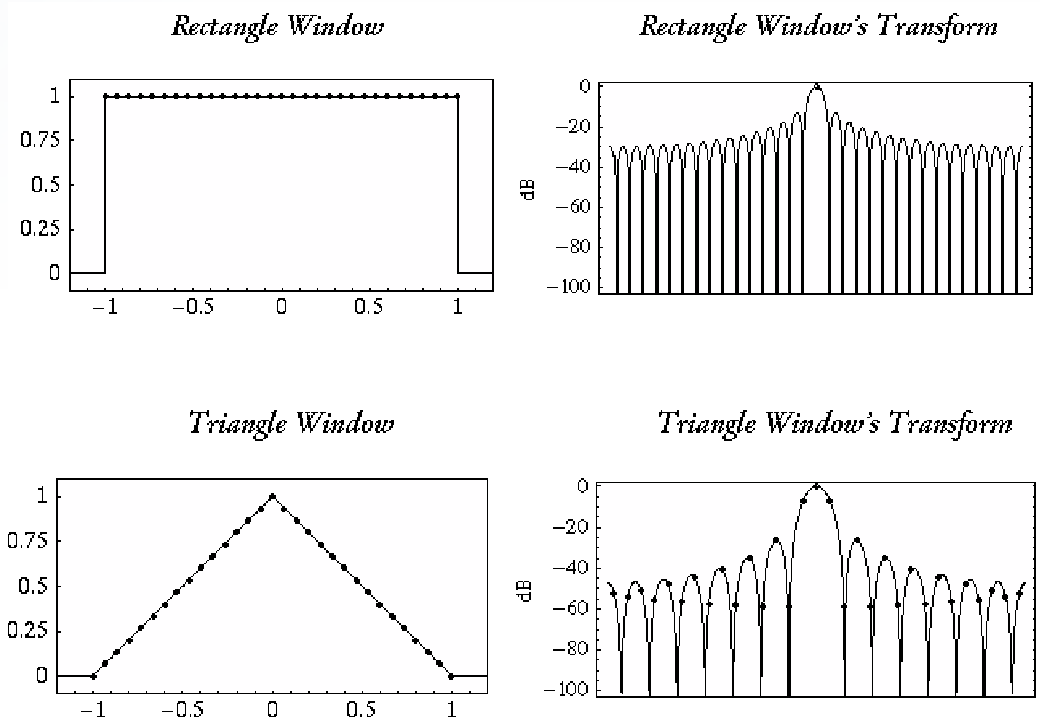
Speech Spectrogram

- Kinds of spectral plots
- Features

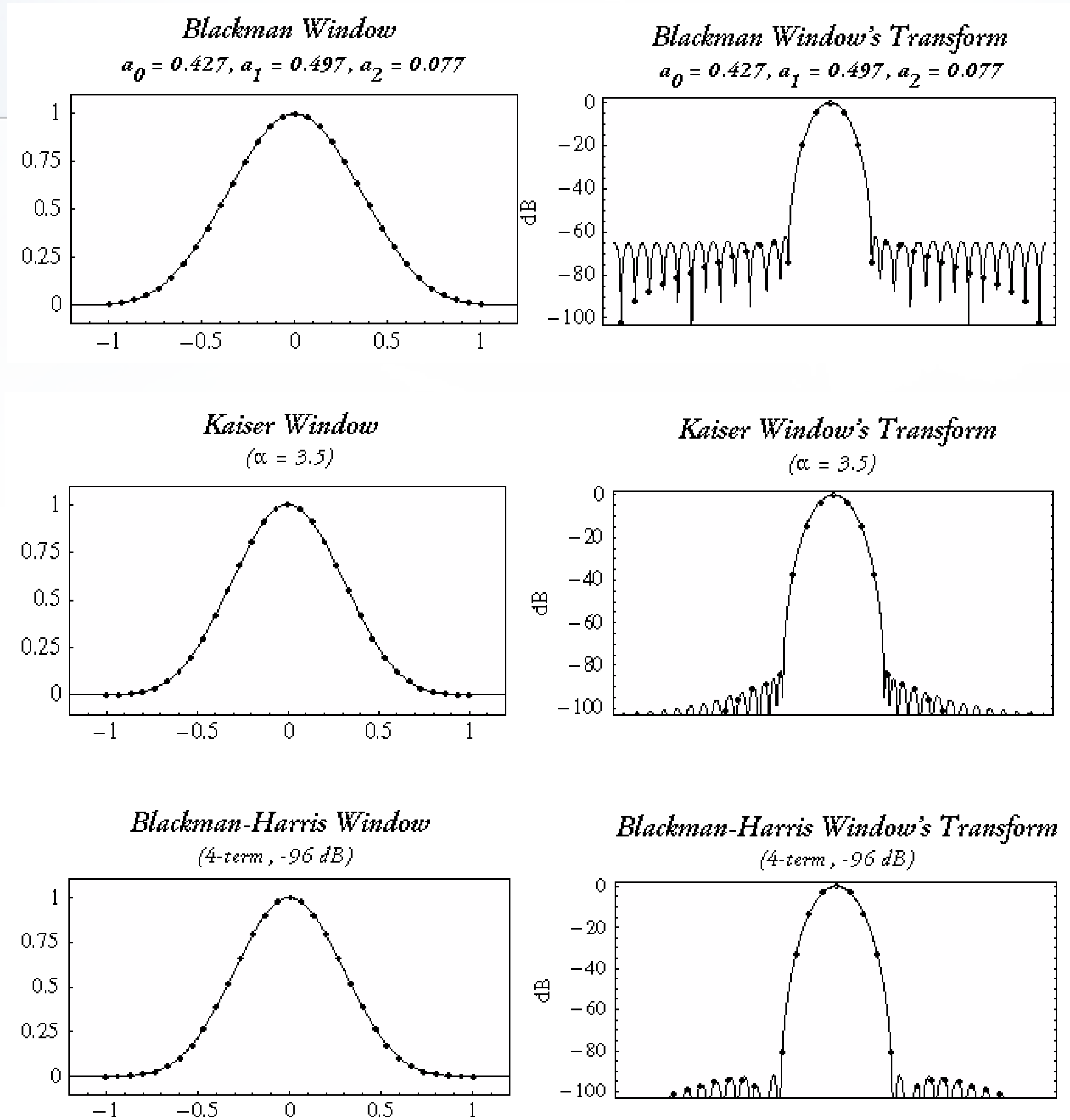


Windows and their Spectra

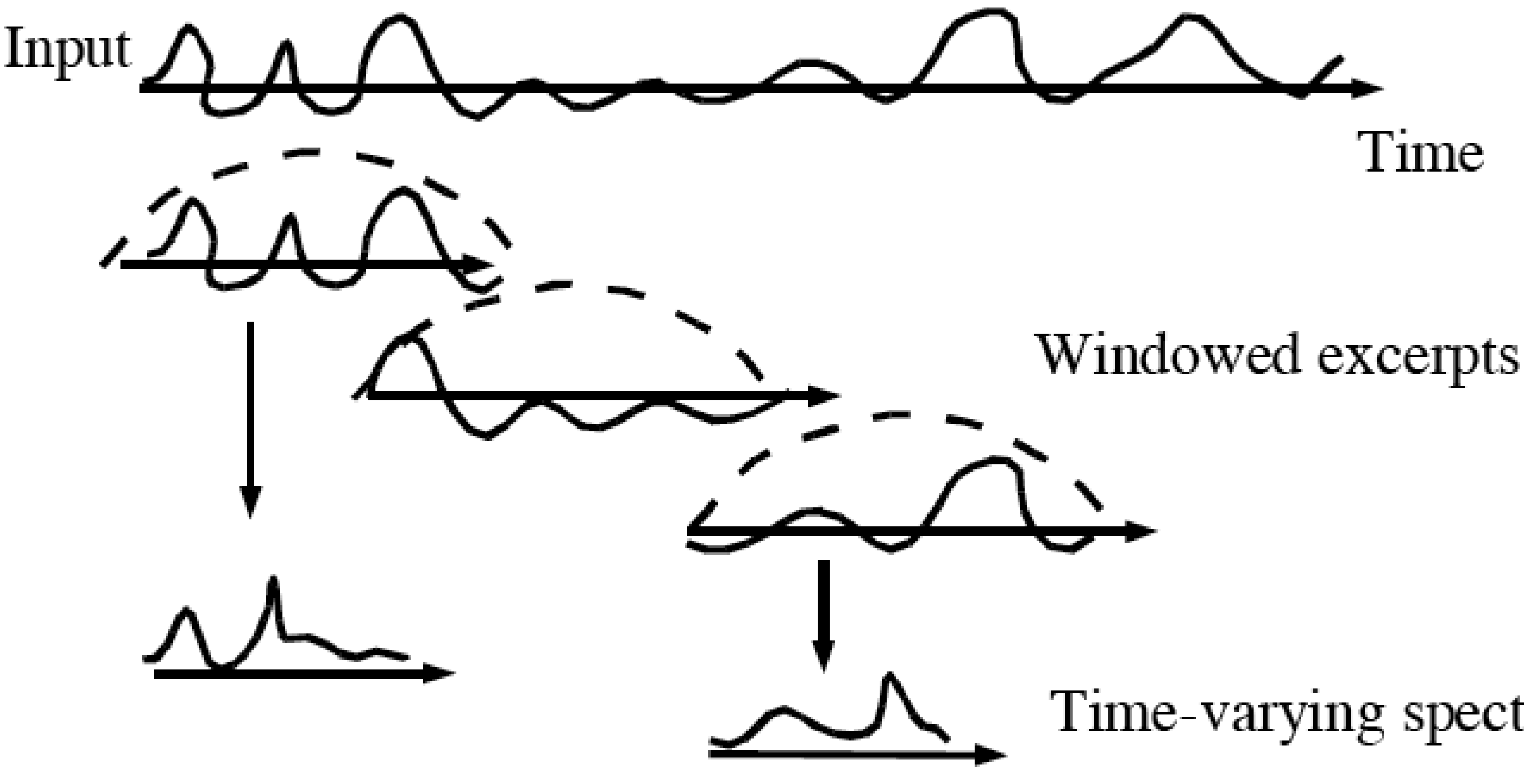
- Trade-offs between window characteristics
- Different windows for different analysis domains



Advanced Windows for Spectral Analysis

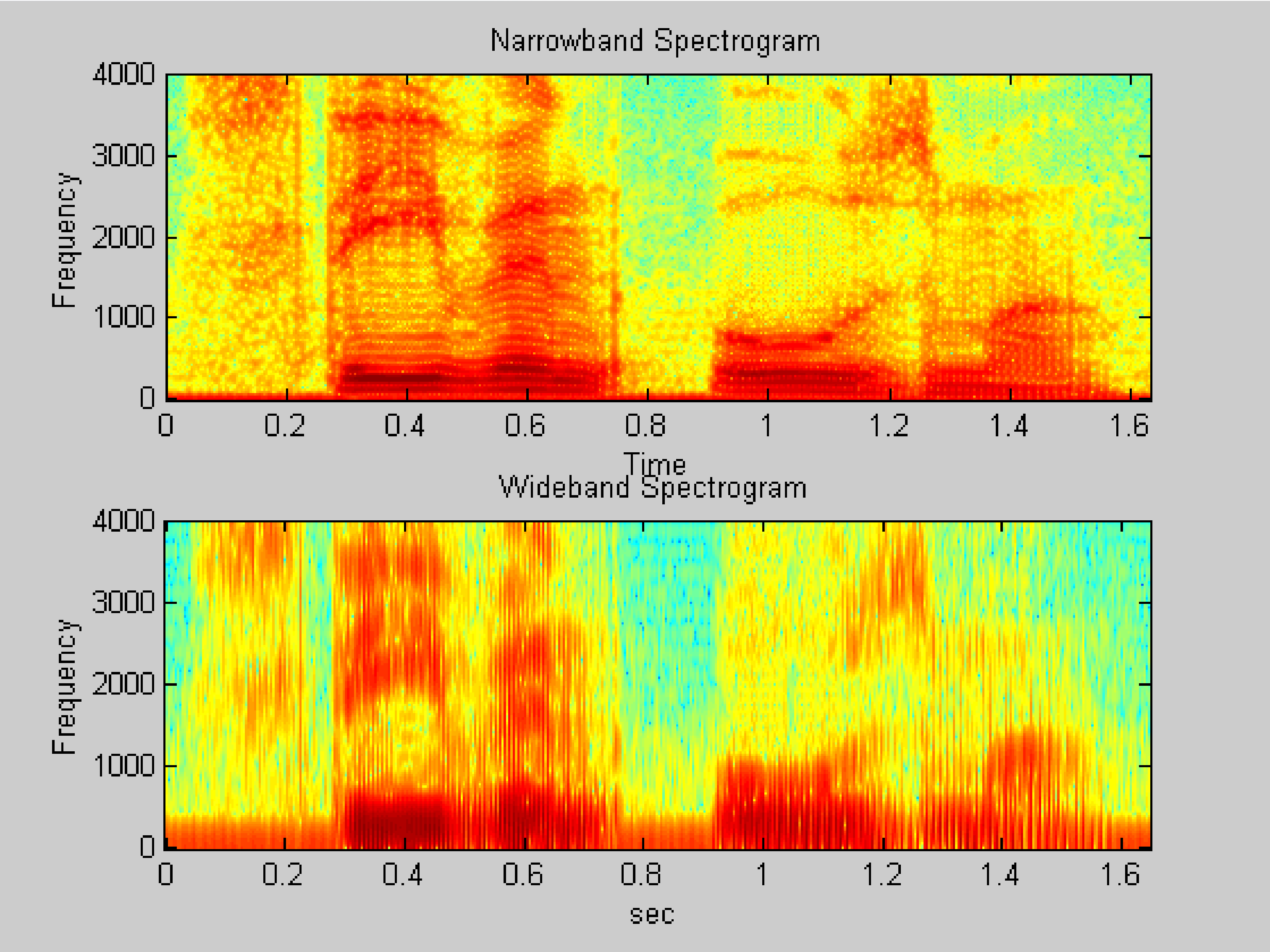


Windowing and the STFT



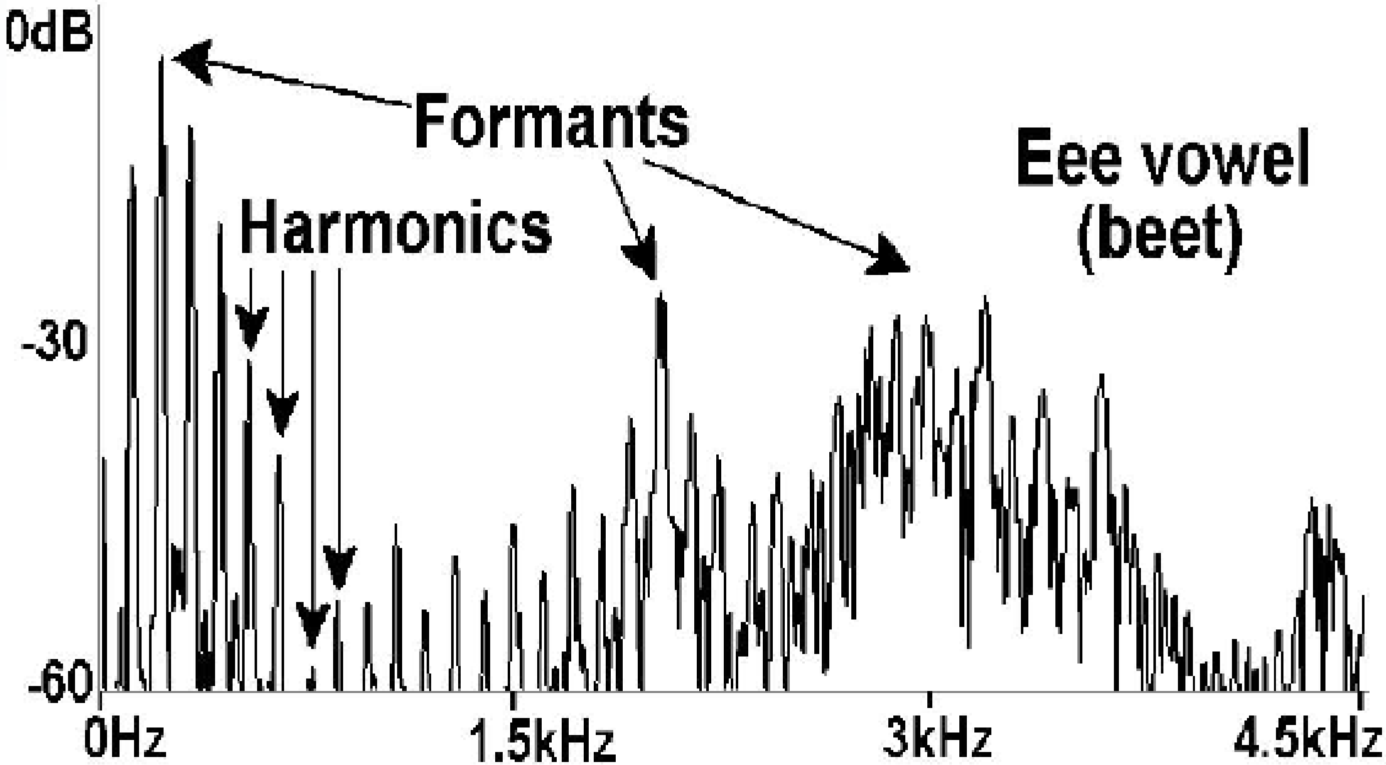
Fast Fourier Transform FFT

The Pitch/Time Trade-off

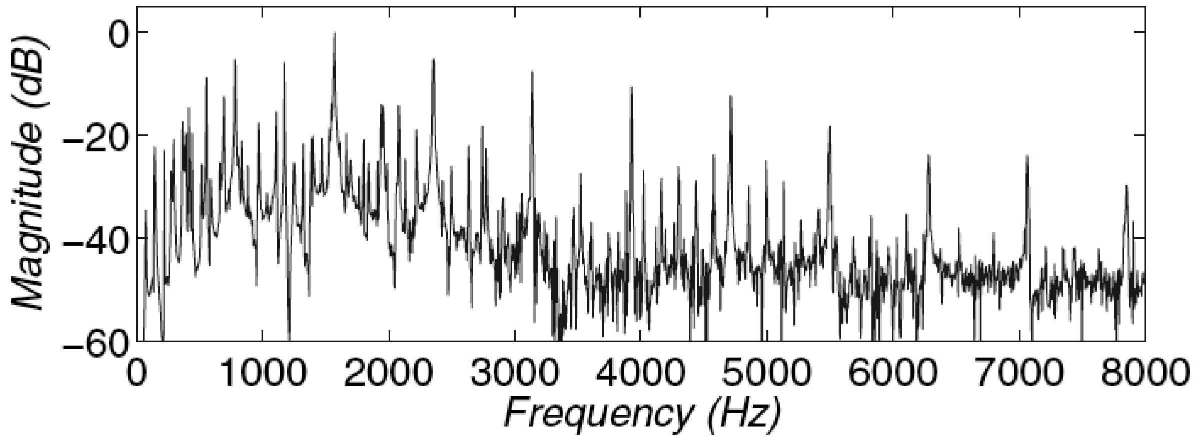


Harmonics and Formants

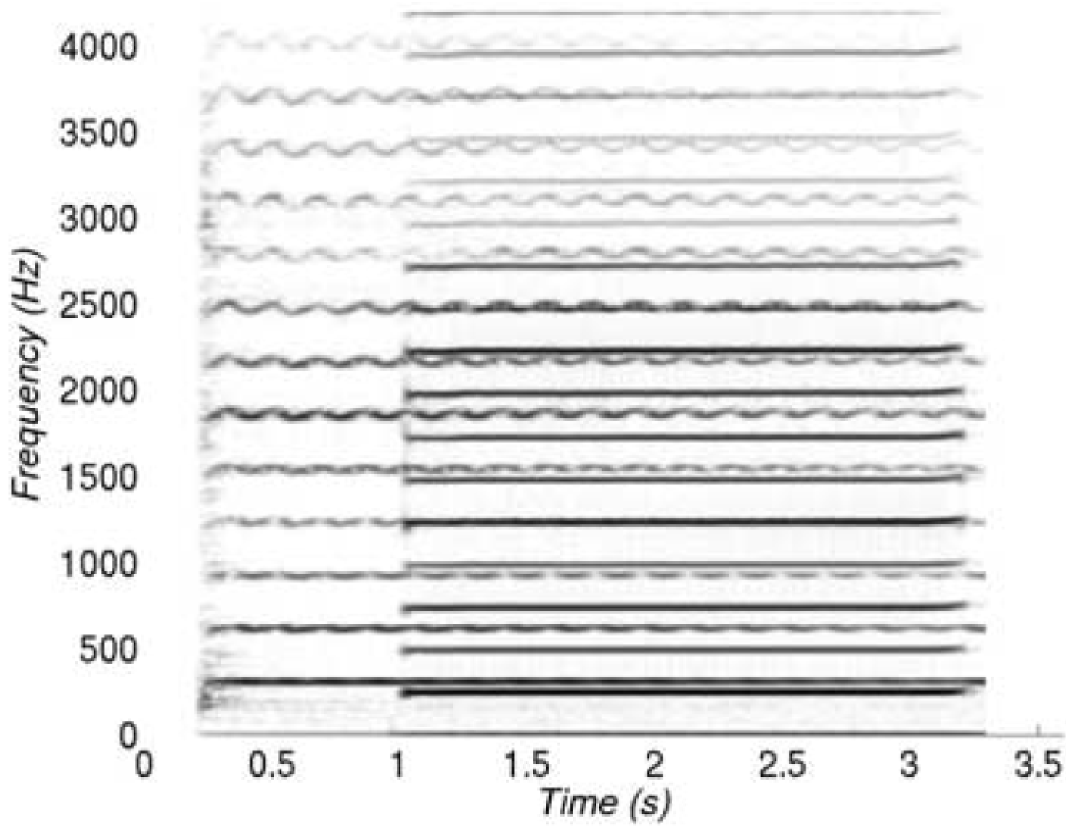
- Source/Filter – instrument resonances



Composite Spectra



- How to disambiguate?
- Track birth/death statistics
- Vibrato (see figure)
- Statistical techniques



Spectral Analysis and Interpretation

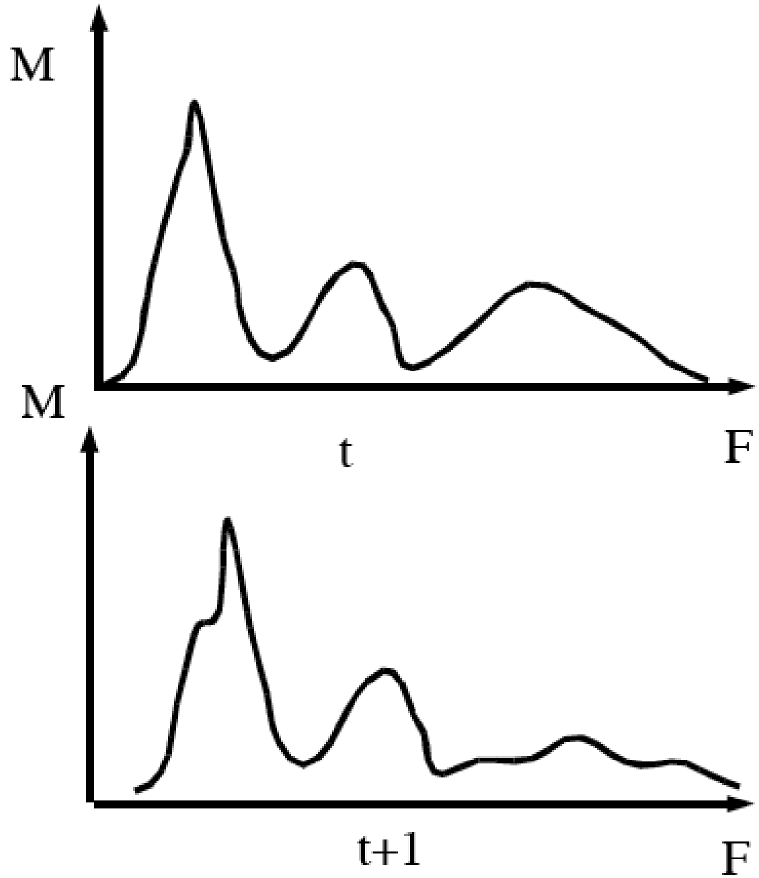
- Spectral data extraction
  - Base frequency
  - Overtone spectrum
  - Formants, resonances, regions
  - Instrument signatures
- Spectral statistics
  - Peak, mean, average, centroid, slope, etc.
  - Spectral variety, etc.

Spectral Moments

- Spectral Centroid – (Mean) 1st moment
- Variance – 2nd moment
- Skew – 3rd Moment
- Kurtosis 4th Moment

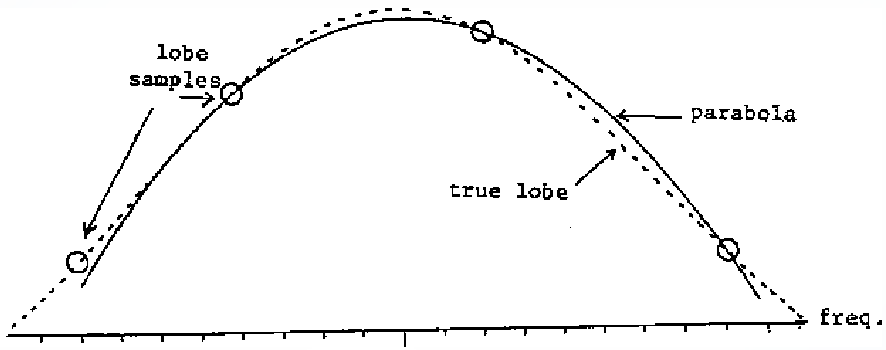
Spectra as Time-varying

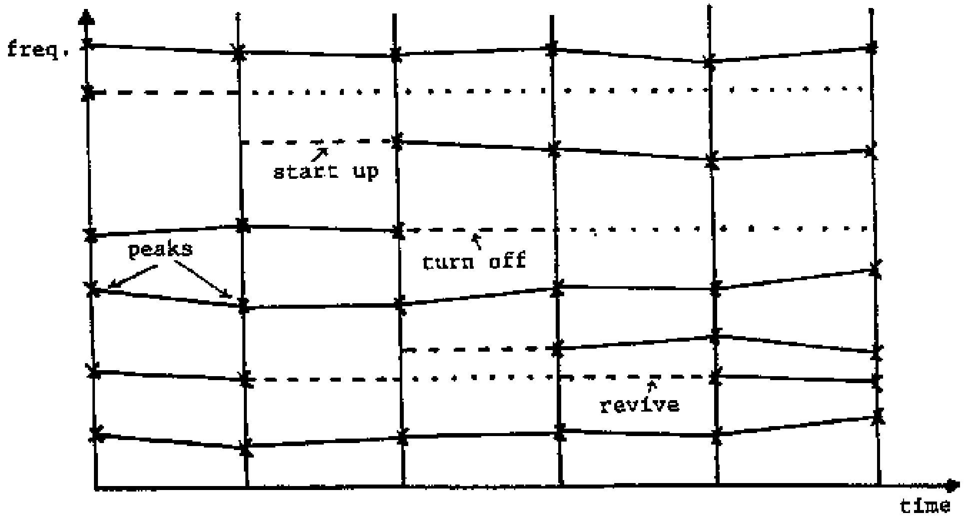
- Track peaks/regions between frames (requires thresholds of change)
- Model the dynamicity (e.g., formant trajectory, vibrato extraction)



Spectral Peaks and Tracking

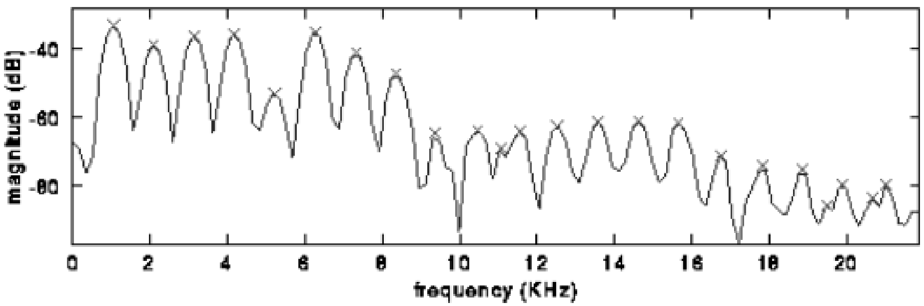
- Peak finding (remember autocorrelation?)
- Peak discrimination
- Peak continuation: tracks and guides
- Derived statistics
- Problem cases

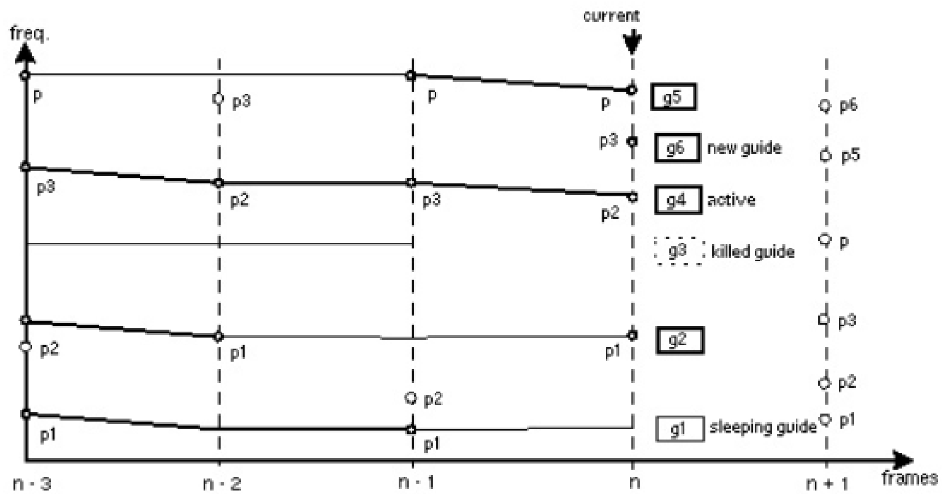




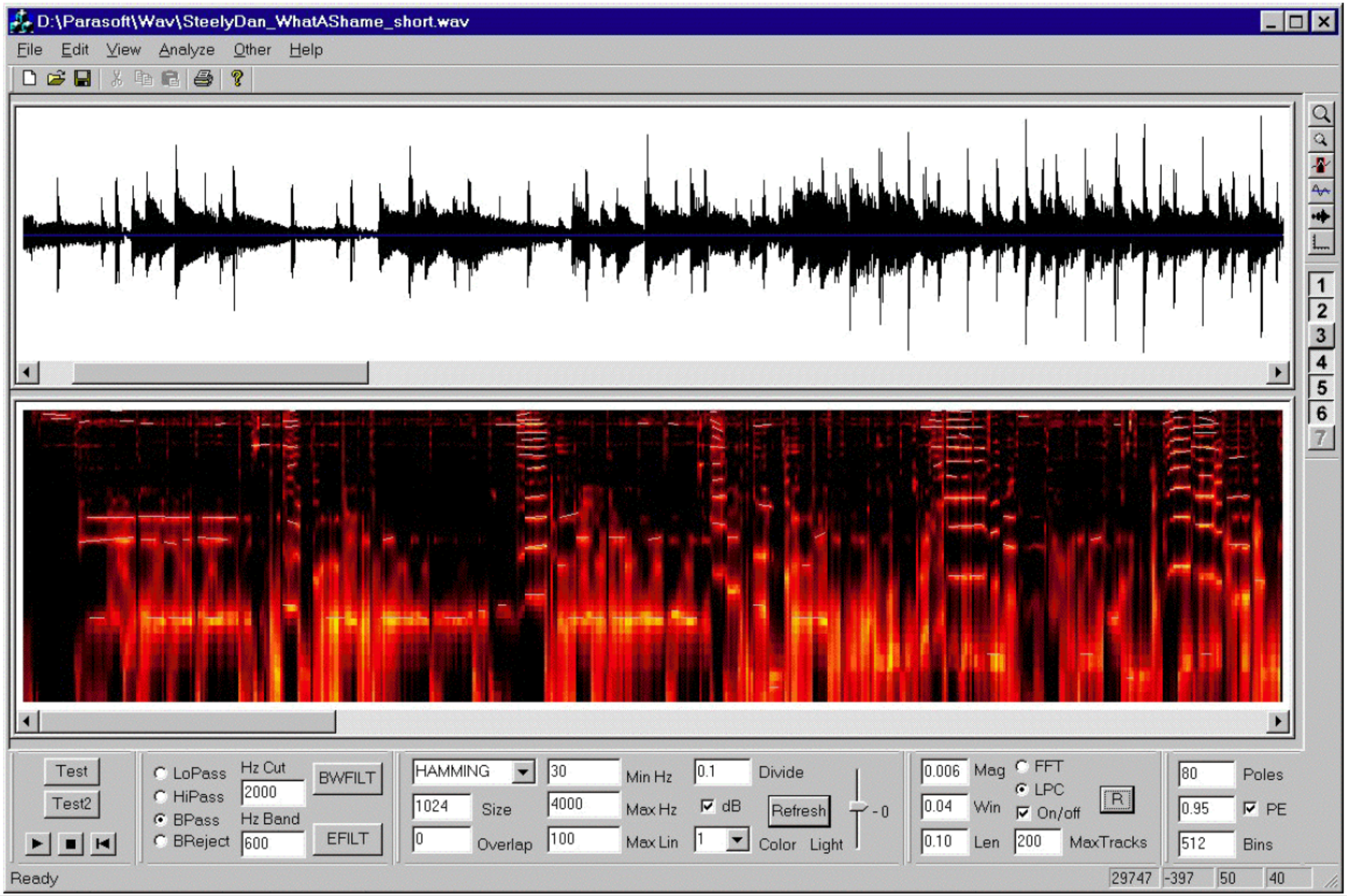
Peaks and Tracks

- Peak-finding
  - Thresholds, distances, heuristics
- Peak-continuation
  - Inter-frame distances and guides
  - Dropped frames and stretching
  - Track birth/death criteria



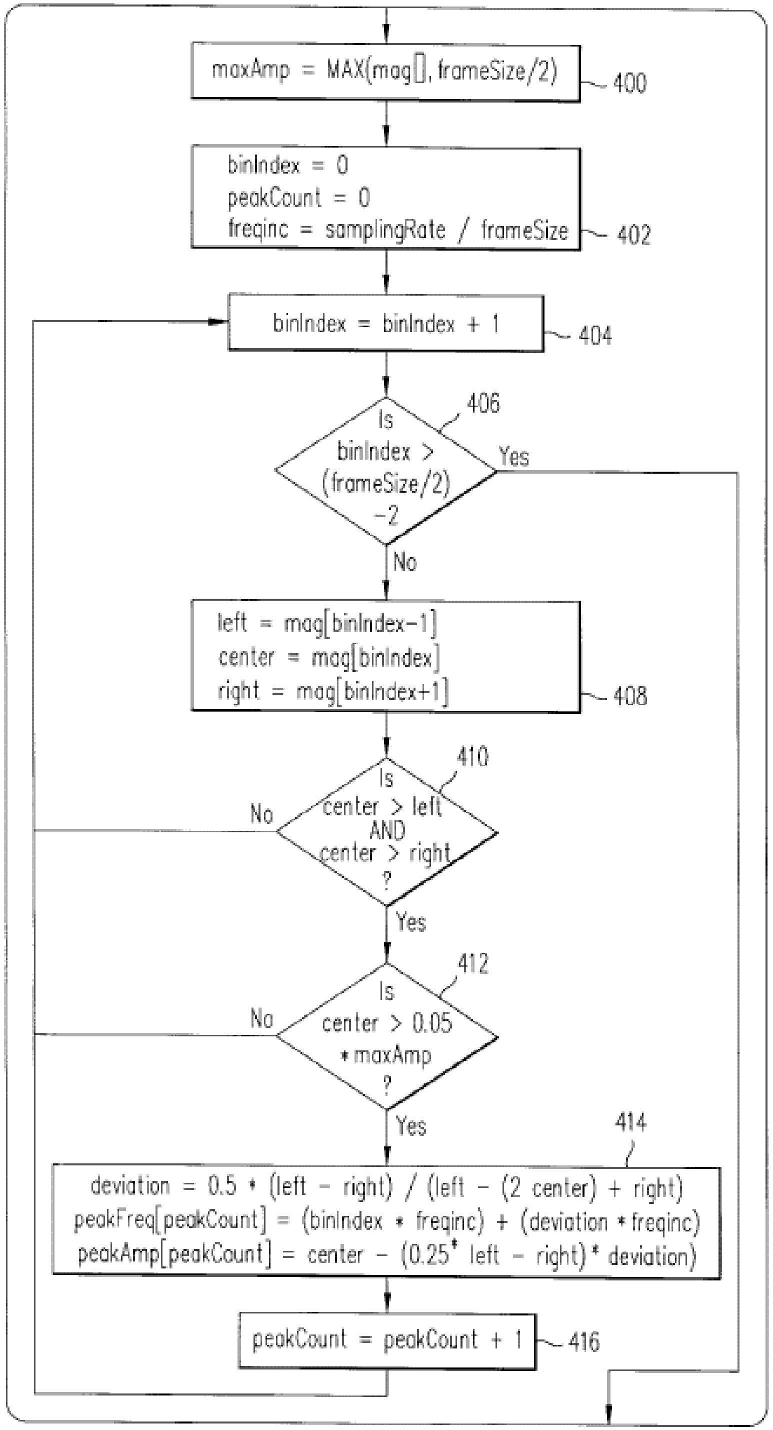


Spectral Peak-Tracking Example



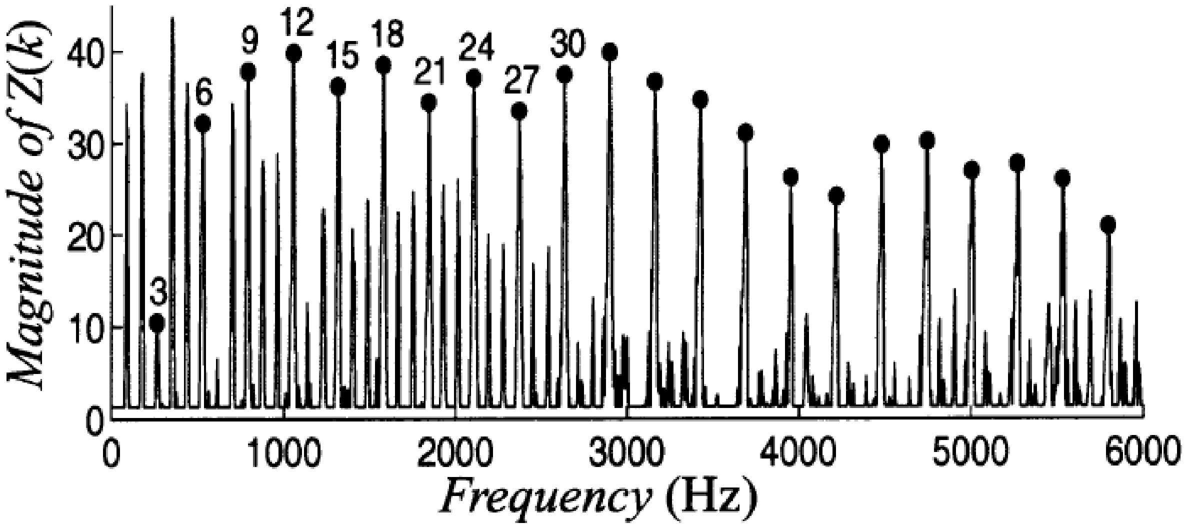
Spectral Peak Detection Algorithm

- From Blum et al. patent # 5,918,223

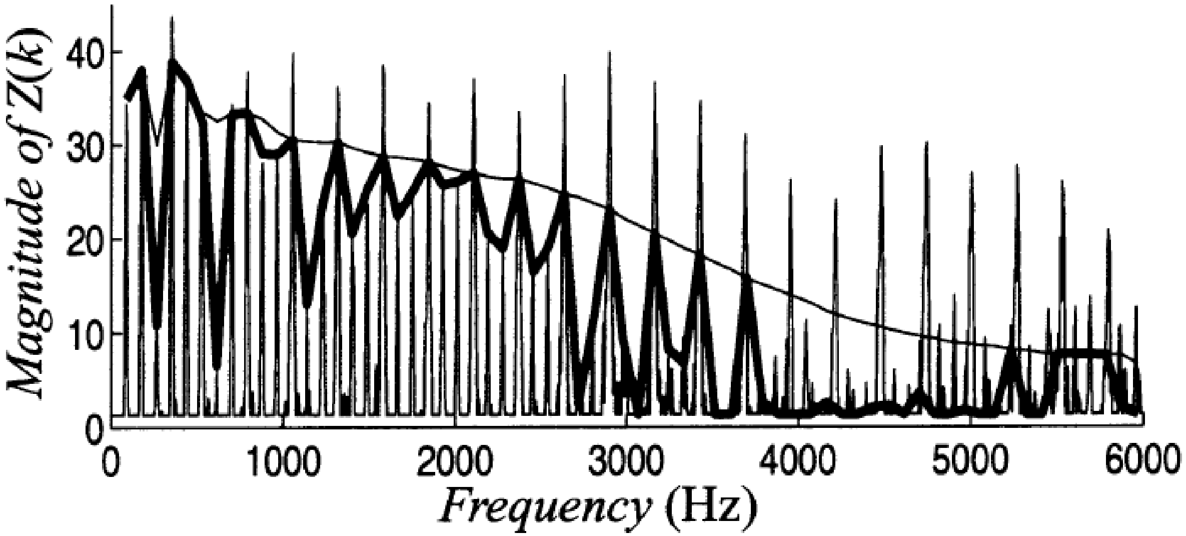


Monday, June 25, 2012

Spectral Smoothness Measure

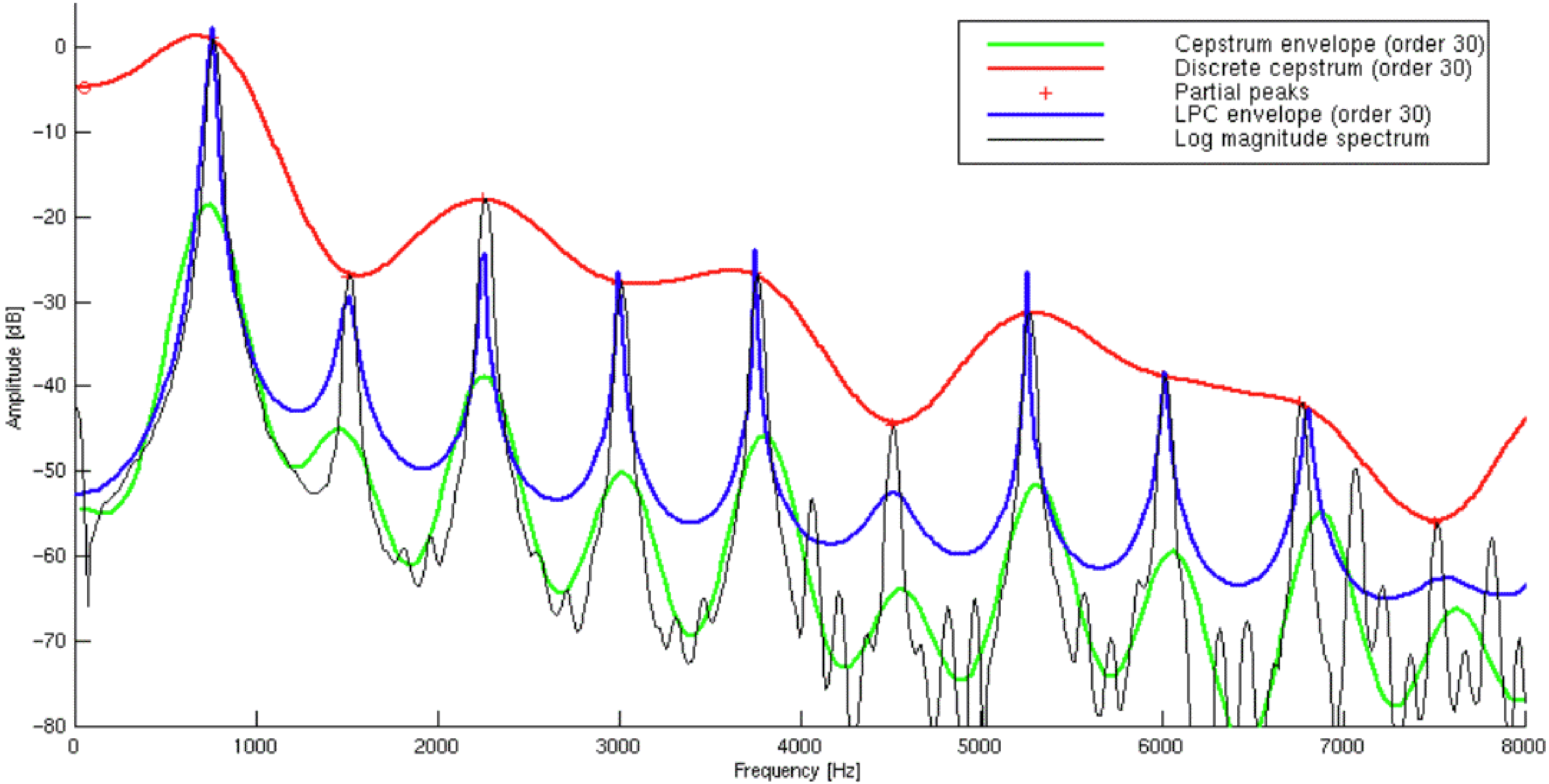


(a)

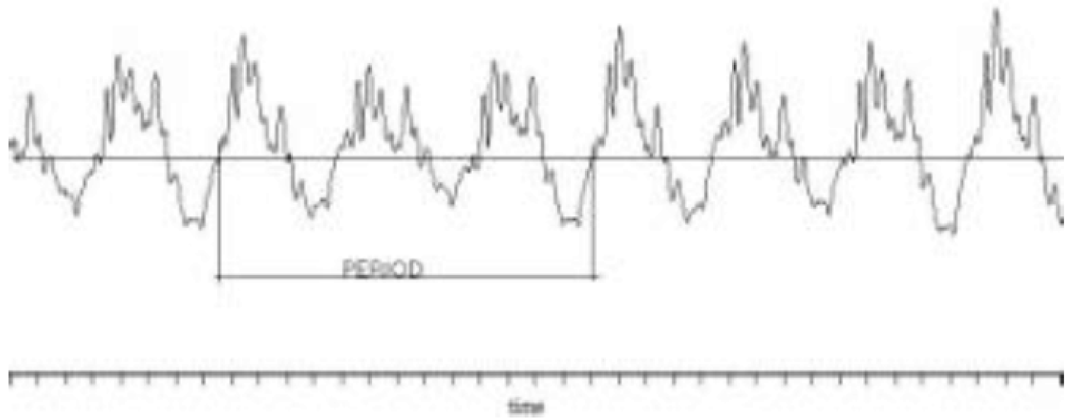


(b)

Smoothed Spectrum Types

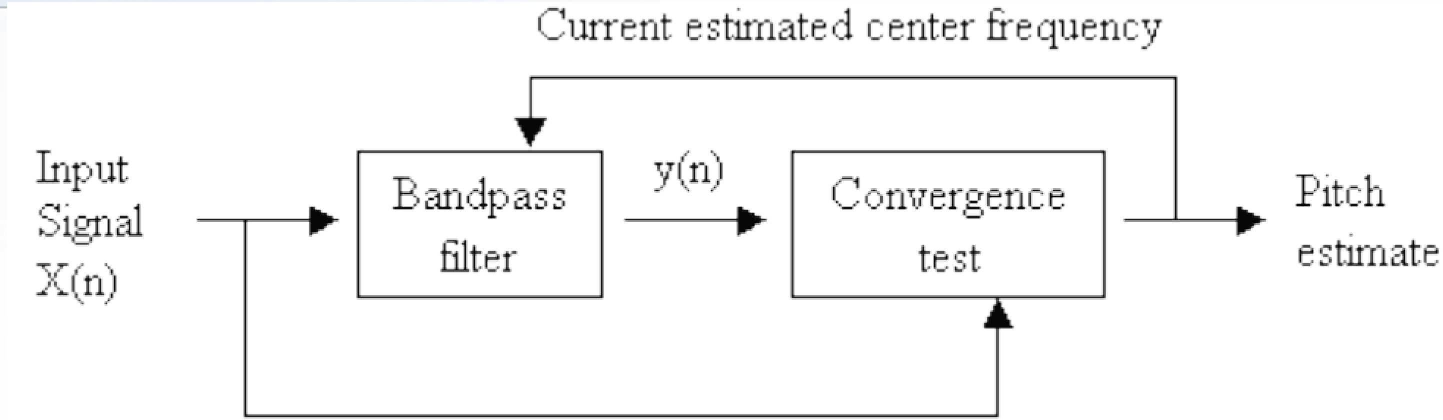


Pitch Detection Techniques



- Find the period of a "periodic" signal
  - First guess whether or not it's periodic
- Simple techniques work for many signals
  - Zero-crossings (with direction, slope)
  - Autocorrelation (with range limitation)
- But it's hard to tell when they fail
  - Random data, silence
  - Octave over/under-tone errors

Filter-based Pitch Detection

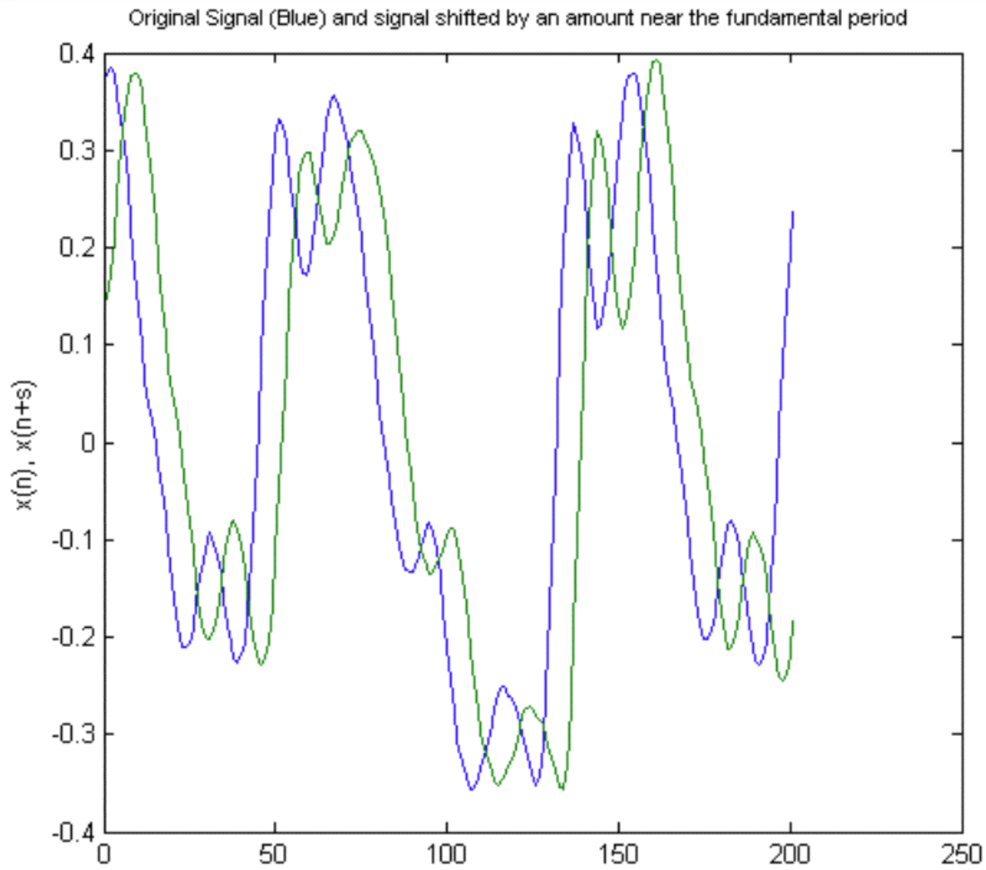


Simple adaptive process for single-frequency source with strong fundamental (i.e., many, but not all, instruments and voices)

- Easily implemented in analog circuitry
- Many variations

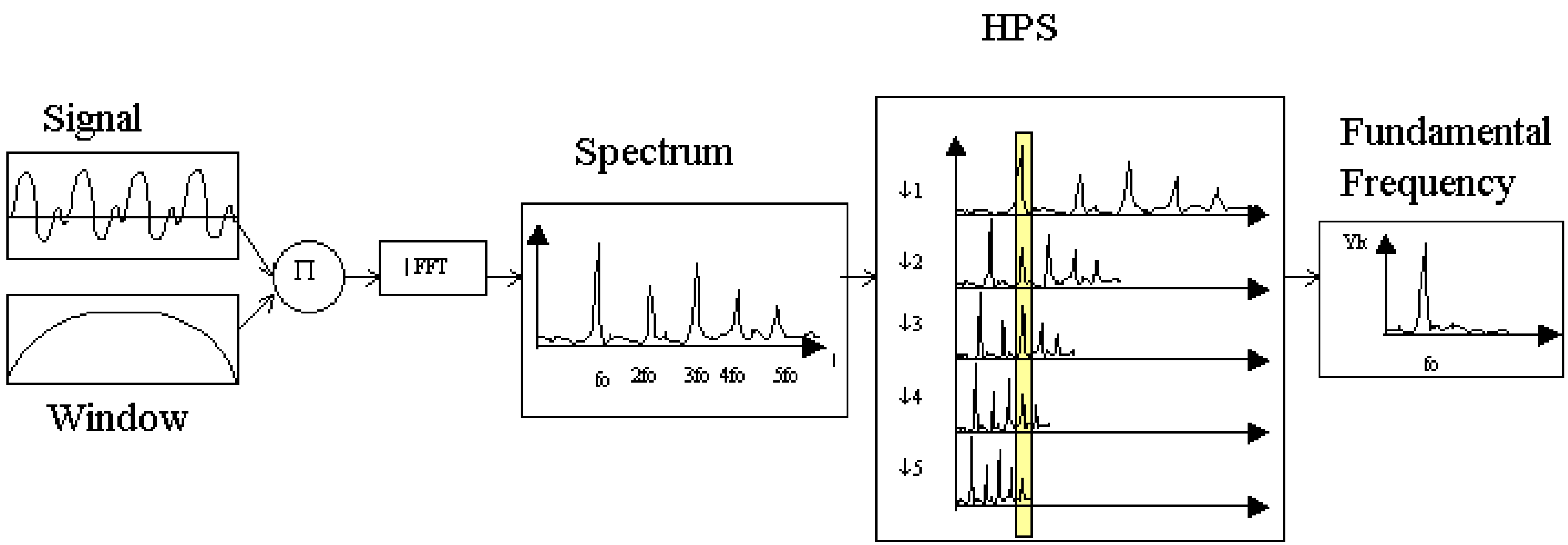
Auto-Correlation

- Slide a signal across itself, taking the vector product at each step
- This AC array has a peak at 0, and the period of the signal
- No peaks for noise



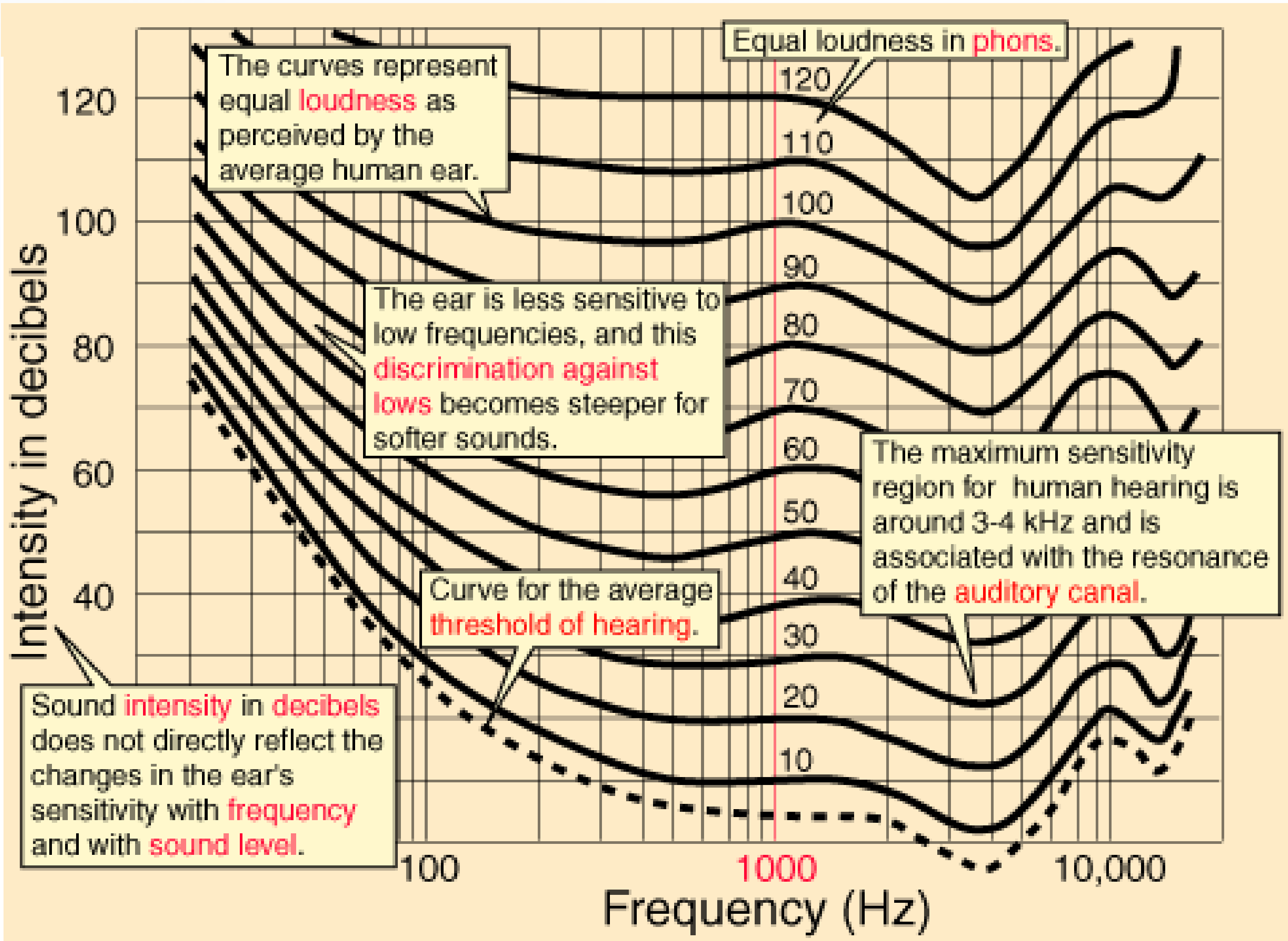
Harmonic Product Spectra

- Decimation of FFT spectra, summation, and spectral peak location
- Assumes overtones are significant, not that fundamental is



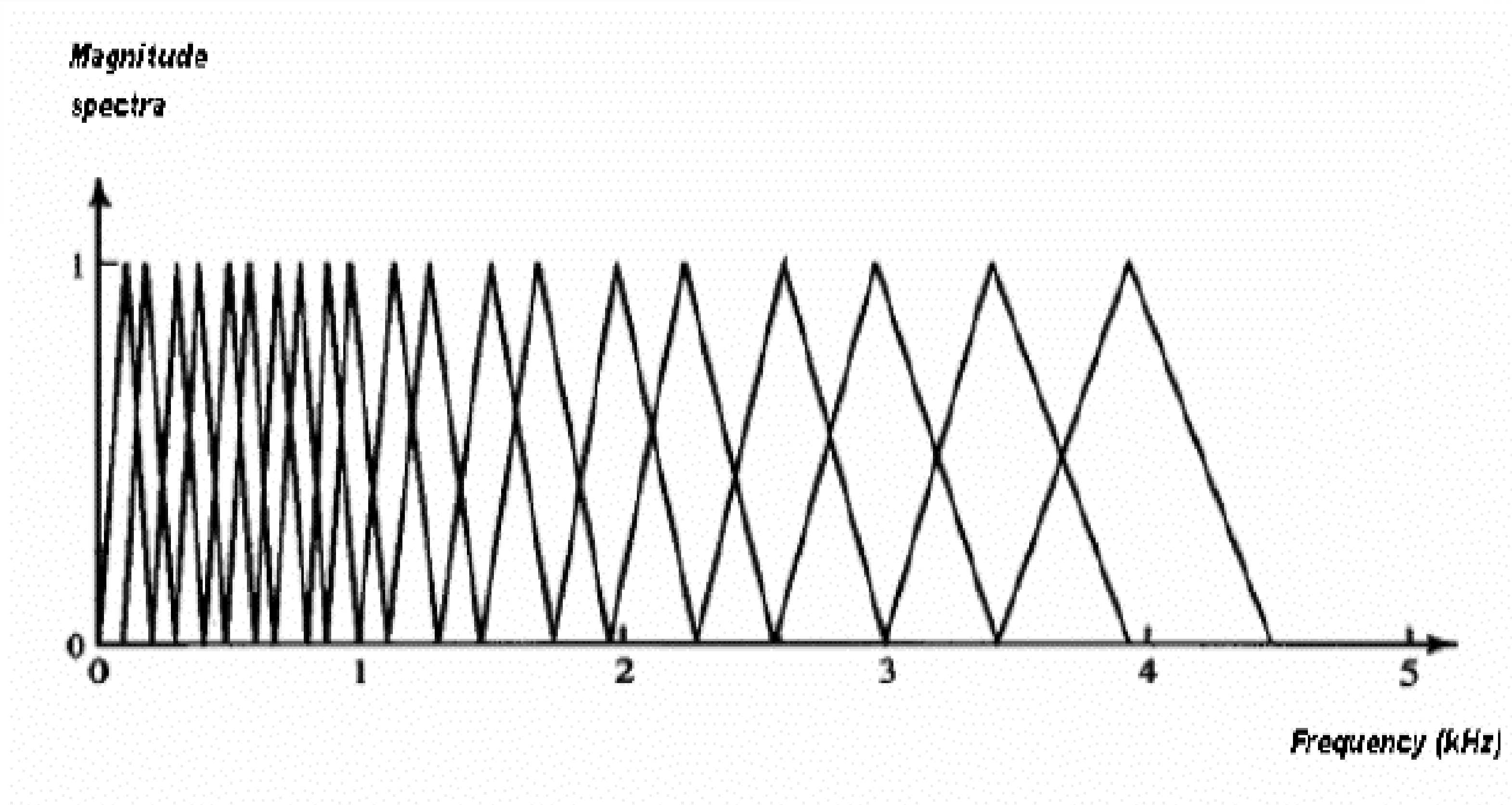
Equal-loudness Curves

- Fletcher-Munson vs. Robinson-Dadson



Frequency Regions and Scaling

- Mel-warped frequency bands



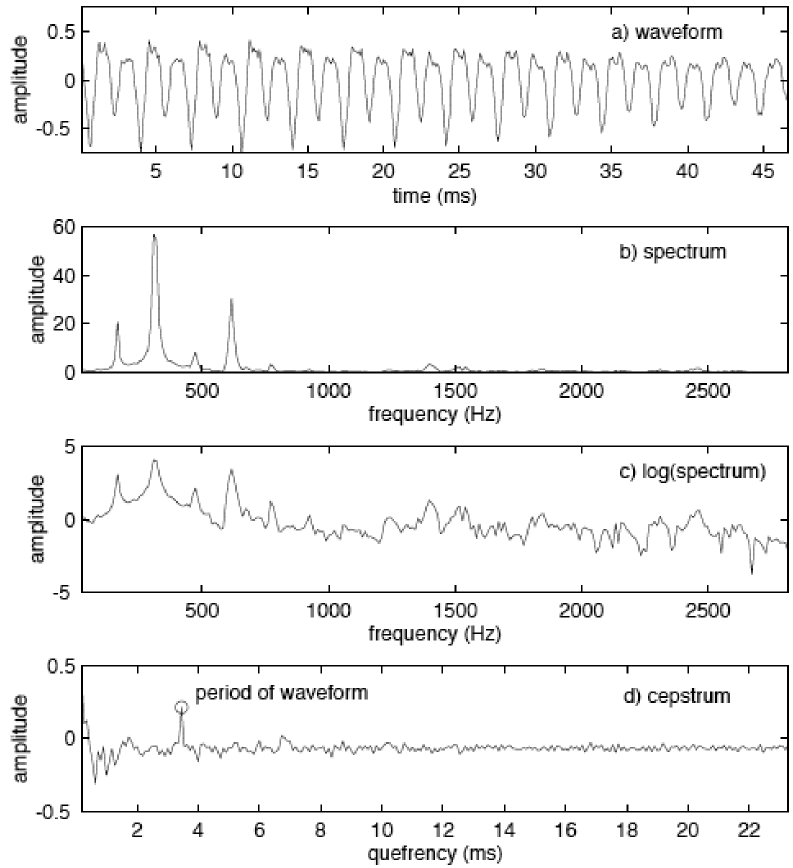
Mel-Freq Cepstral Coefficients

- Steps:
  - Signal
  - FT
  - Log magnitude
  - Phase unwrapping
  - FT (or DCT)
- Interpretations
  - “ceps” = “Spec”trum of spectrum
  - “Quefrency”
  - Mel-scale
  - Mel-scale filters

- Instead of AC, use FFT or DCT of PDS
- Leads to interesting statistics of higher-level spectral properties, see next section

MFCC Analysis

- Analogy
  - Start with log spectrum of mixed complex tones: several sets of related partial peaks
  - Take, e.g., the autocorr. of the FFT PDS
  - Warped frequencies of peaks correspond to fundamental frequencies of overtone series

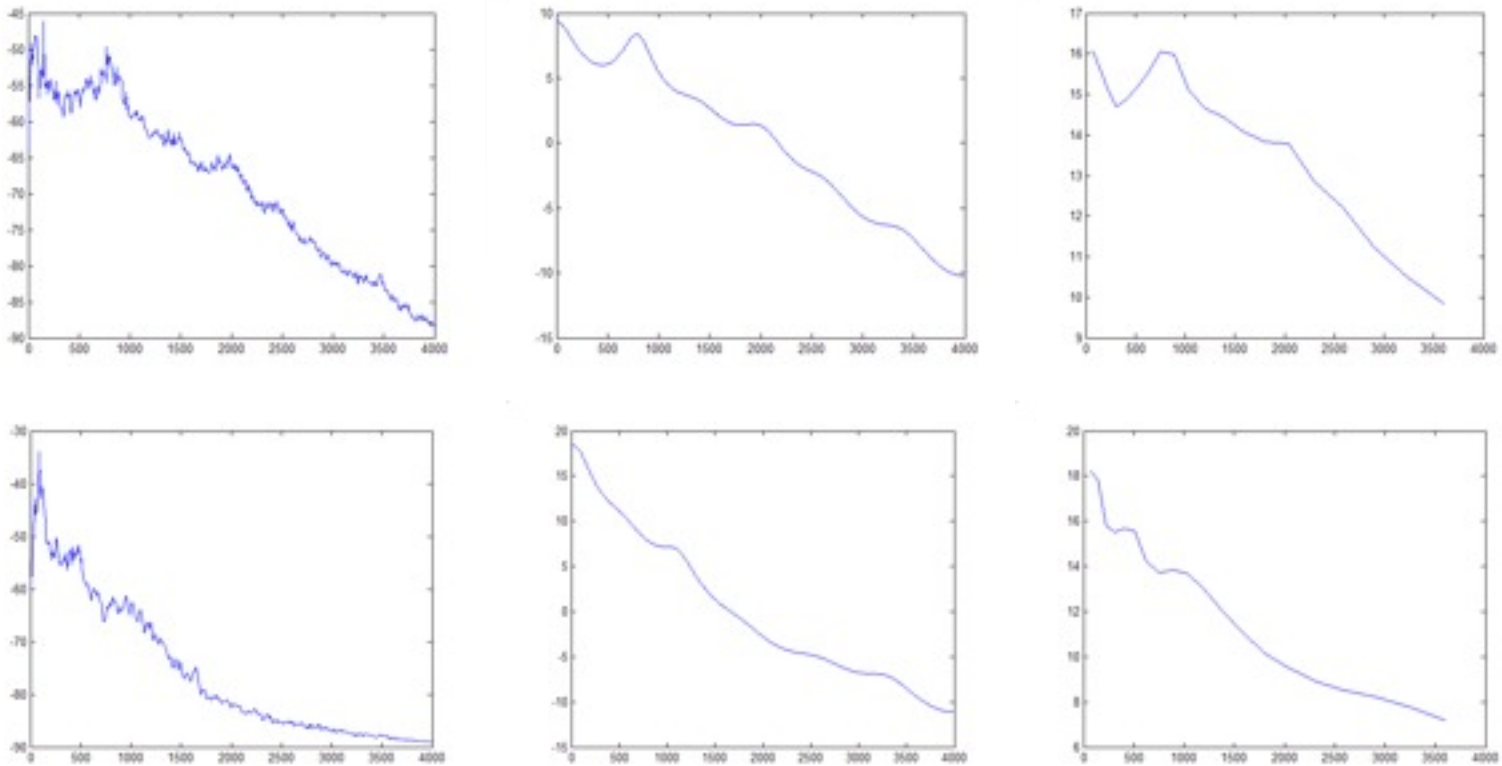


Comparison With LPC (by Andrianakis & White)

FFT PSD | 13-pole LPC Spectrum | 20 Mel Spectrum

Train

Car



Frequency [Hz]

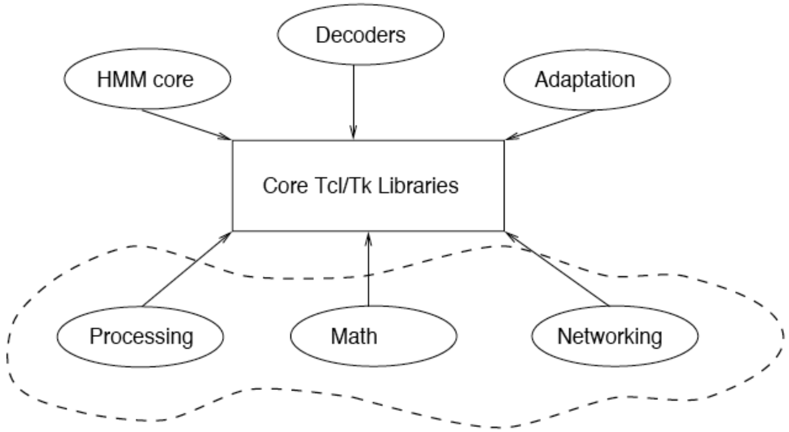
Spatial-domain Features

- M/S Encoding (stereo sum & difference)
- Surround-sound processing
  - L/R vs C
  - L/R vs Ls/Rs
- Frequency-dependent spatial separation
- Higher-dimensional sources
- Stem tracks

Other Feature Domains

- Other time-domain features
  - Beats, beat histograms
- Other frequency-domain features
  - Fluctuation patterns
- Other time-frequency transforms
  - Filter banks
- Wavelets
  - Trades off temporal & spectral resolution
- Linear Predictive Coding
  - polynomial representation

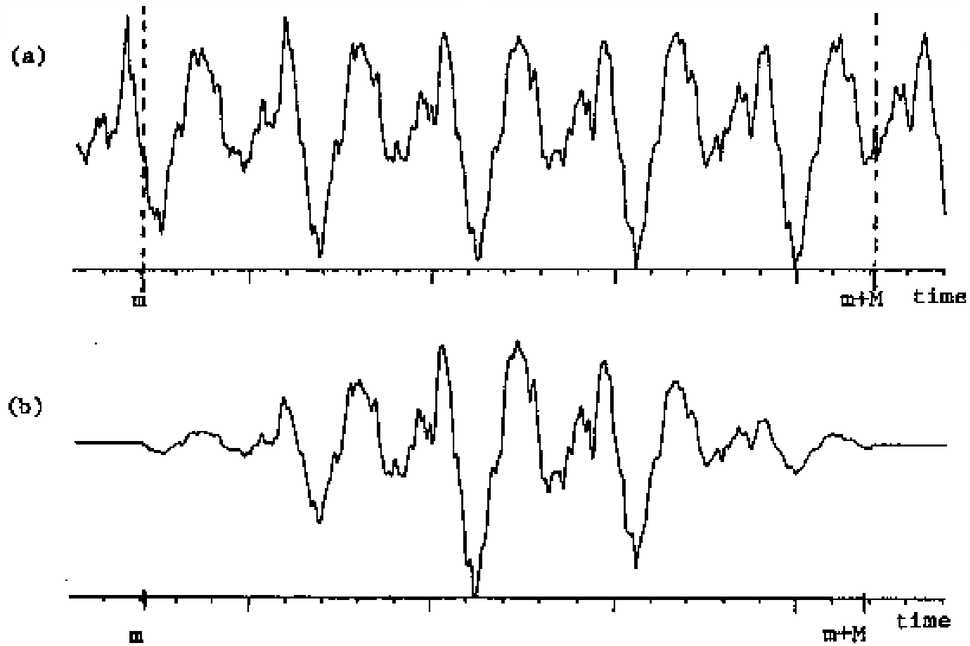
APIs for MIR Tools



- Marsyas: G. Tzanetakis (11), flexible tool set, scripting language, segmentation and classification
- LibOFA: Holm/Pope (00), simple FV for unique ID comparing to a large pre-analyzed database
- D2K/M2K: West/MIREX (06), Java-based GUI related to D2K, many apps.
- LibTSP: P. Kabal (00), C routines for DASP & IO
- CSL: STP/MAT (05), C++ class library for DASP, synthesis, control, spatialization and MIR

APIs - 2

- Libxtract
- Aubio
- SonicVisualizer plug-ins
- Loris
- SPEAR
- CSL
- LibTSP



Spectral Tools

- SPEAR
- Loris
- Marsyas
- Sonic visualizer

Review

- MIR Apps
- Signal analysis processing chains
- Feature vector design from app requirements
- Kinds of audio features
- Basic feature statistics