

---

# Music 209

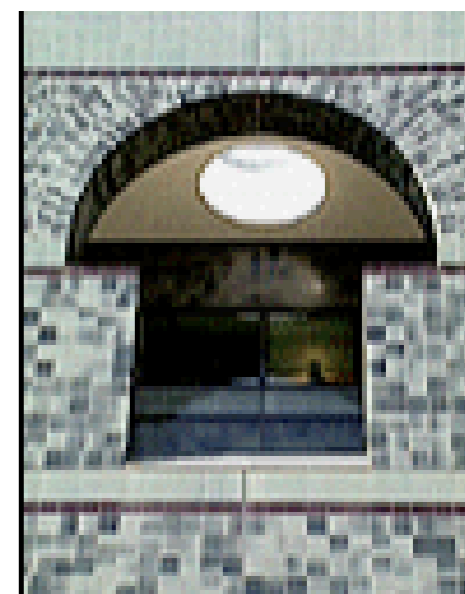
## Advanced Topics in Computer Music

### Lecture 9 – Spectral Methods

---



2006-3-16



**Professor David Wessel (with John Lazzaro)**  
([cnmat.berkeley.edu/~wessel](http://cnmat.berkeley.edu/~wessel), [www.cs.berkeley.edu/~lazzaro](http://www.cs.berkeley.edu/~lazzaro))

---

[www.cs.berkeley.edu/~lazzaro/class/music209](http://www.cs.berkeley.edu/~lazzaro/class/music209)

---



Over the past 2 months ...

**Time-Domain**  
Algorithms  
for  
Sound  
Modification  
and Analysis

# Recall: Note-level Time Warping

---

We now look at time warping algorithms for this problem ...

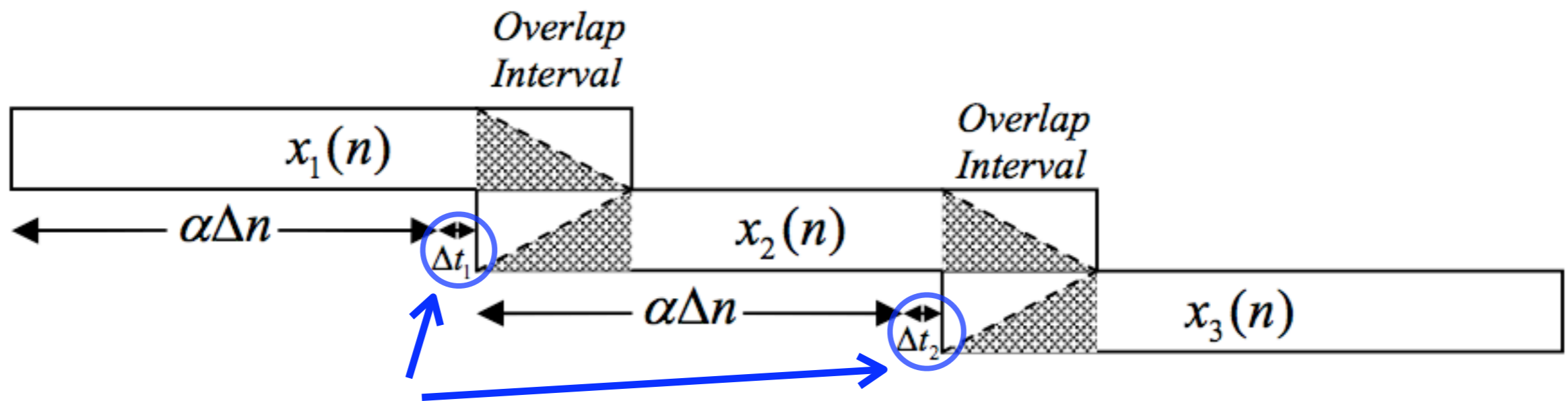
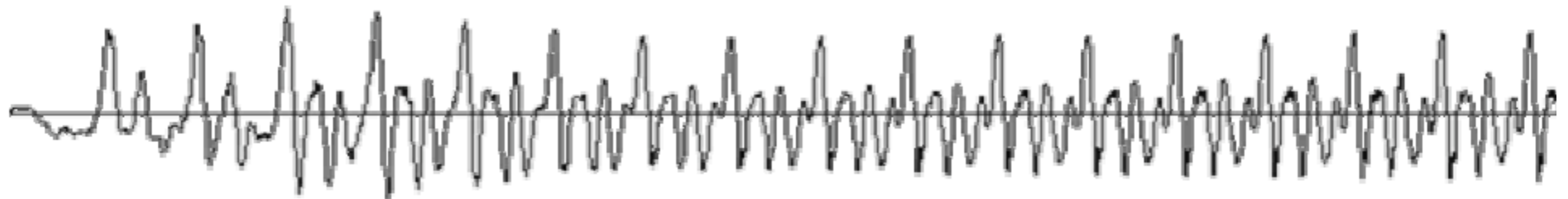


- \* Length of attack transient unchanged.  
Time warp only affects sustained region.
- \* Local temporal properties of sustained region unchanged (example: vibrato speed)
- \* Long-range properties of sustained region stretch or shrink (example: crescendos).

# ← “Synchronous OLA” - SOLA

Recall: Tune each  $\Delta t_i$  to minimize artifacts, and then create final waveform by summing all blocks, doing a crossfade at the overlaps.

← “overlap and add” - OLA



Tuning  $\Delta t_i$  is primary way methods differ.

PSOLA, WSOLA, PICOLA, ...

Another trick: Detect transients, don't OLA them. First use: Lexicon 2400, 1986.

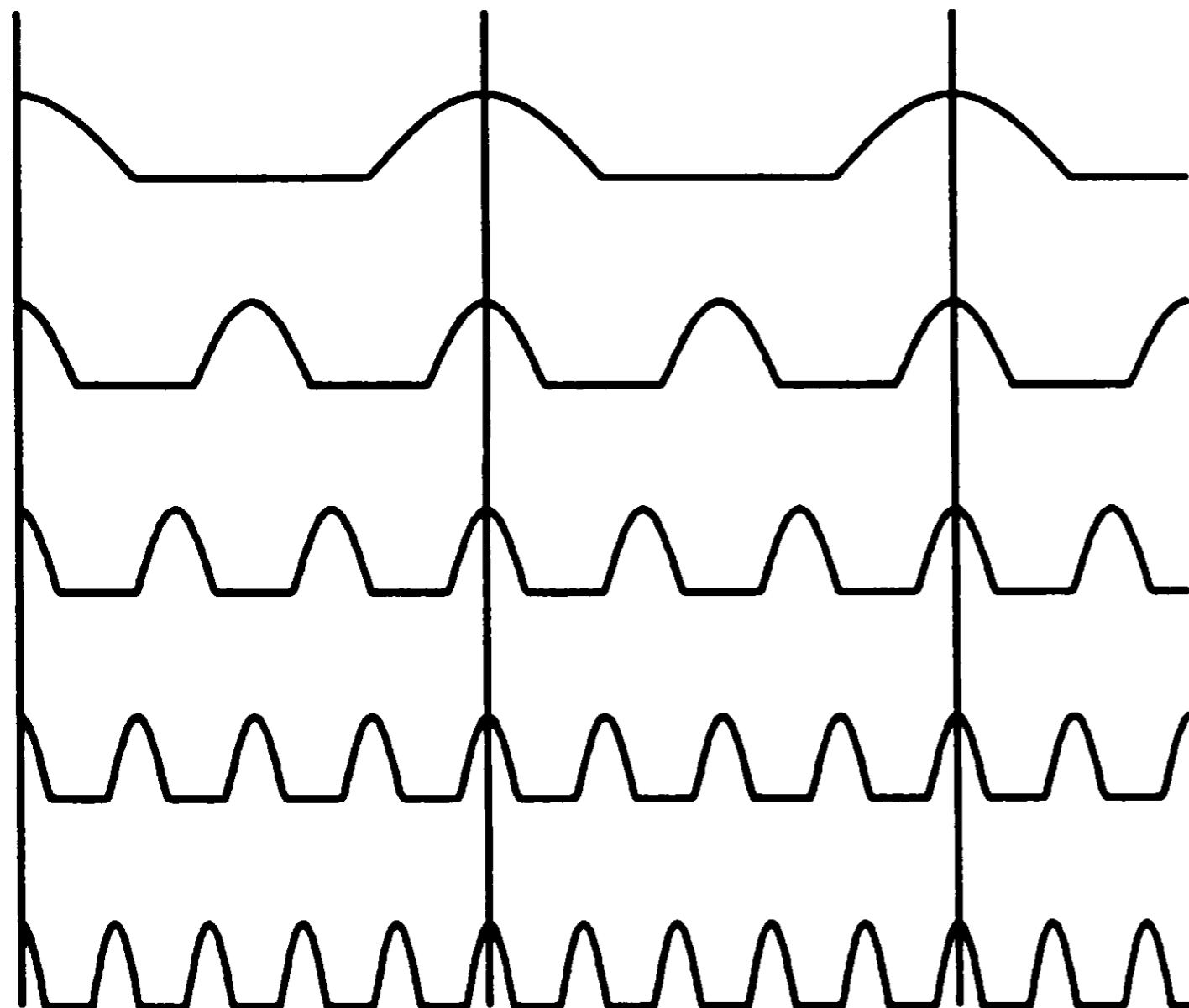
# Recall: Pitch shifting

Original key is F

New key is B flat

The diagram shows three musical staves in 4/4 time. The first staff is in F major (one flat) and contains the notes F, G, A, B. The second staff shows the same notes with arrows indicating a shift of 1, 2, 3, and 4 semitones up. The third staff is in B-flat major (two flats) and contains the notes B-flat, C, D, E-flat, representing a total shift of 5 semitones.

$$+5 \text{ semitones} = ({}^{12}\sqrt{2})^5 = 1.33483985$$



+5 semitones

1.33483985  $f$

$2f$

2.66967971  $f$

$3f$

4.00451956  $f$

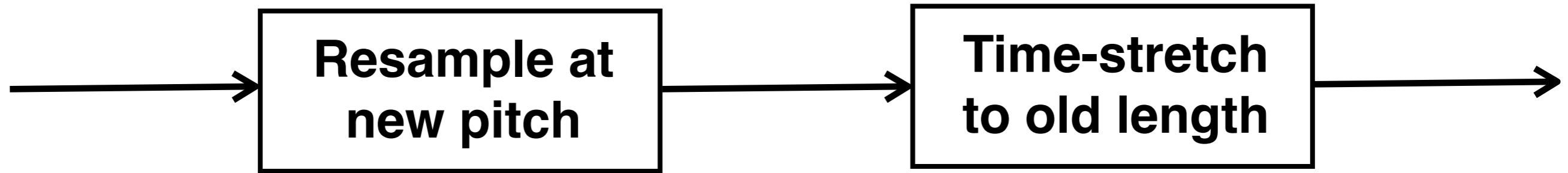
$4f$

5.33935942  $f$

$5f$

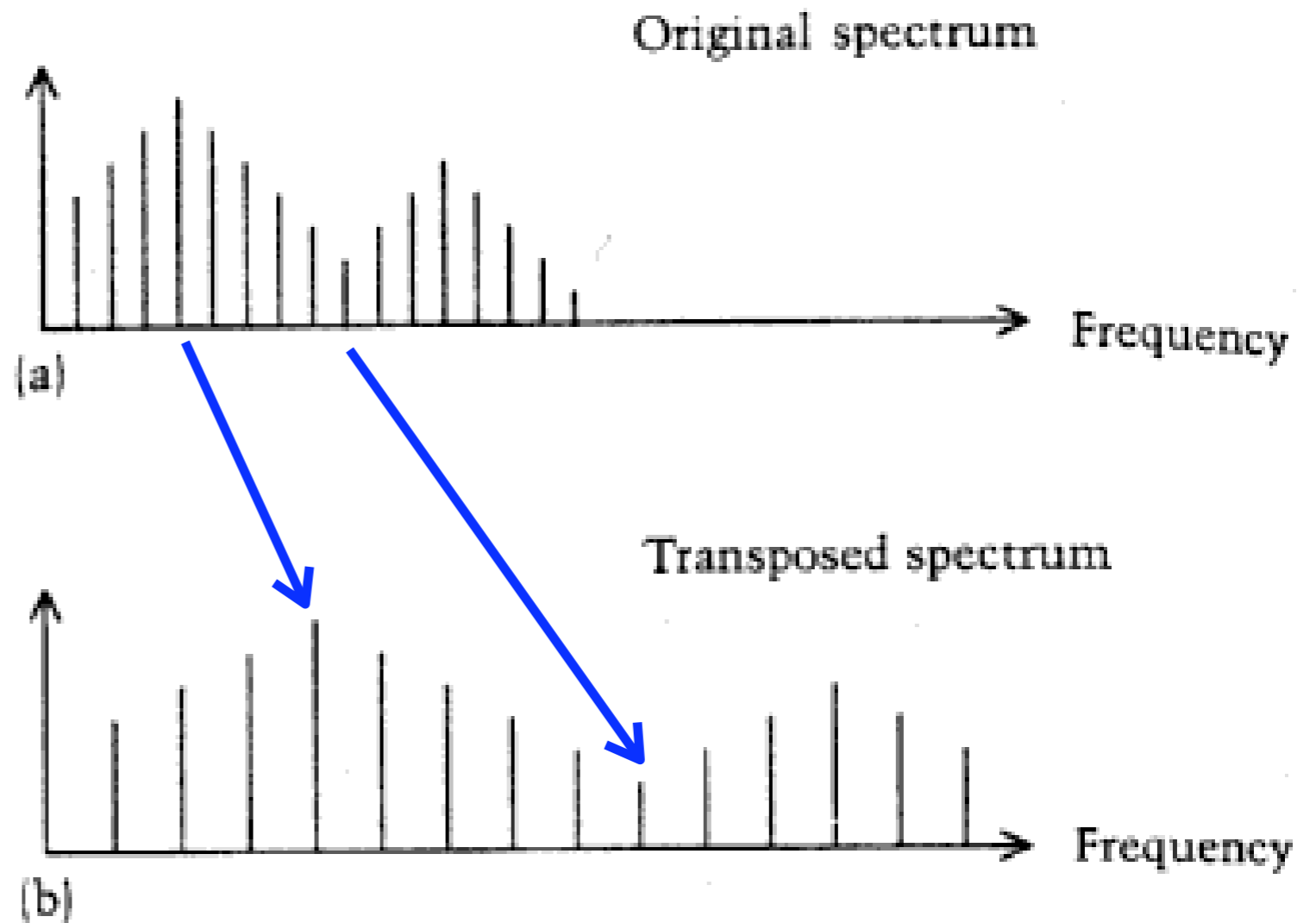
6.67419927  $f$

# Recall: Resample + Time Stretch



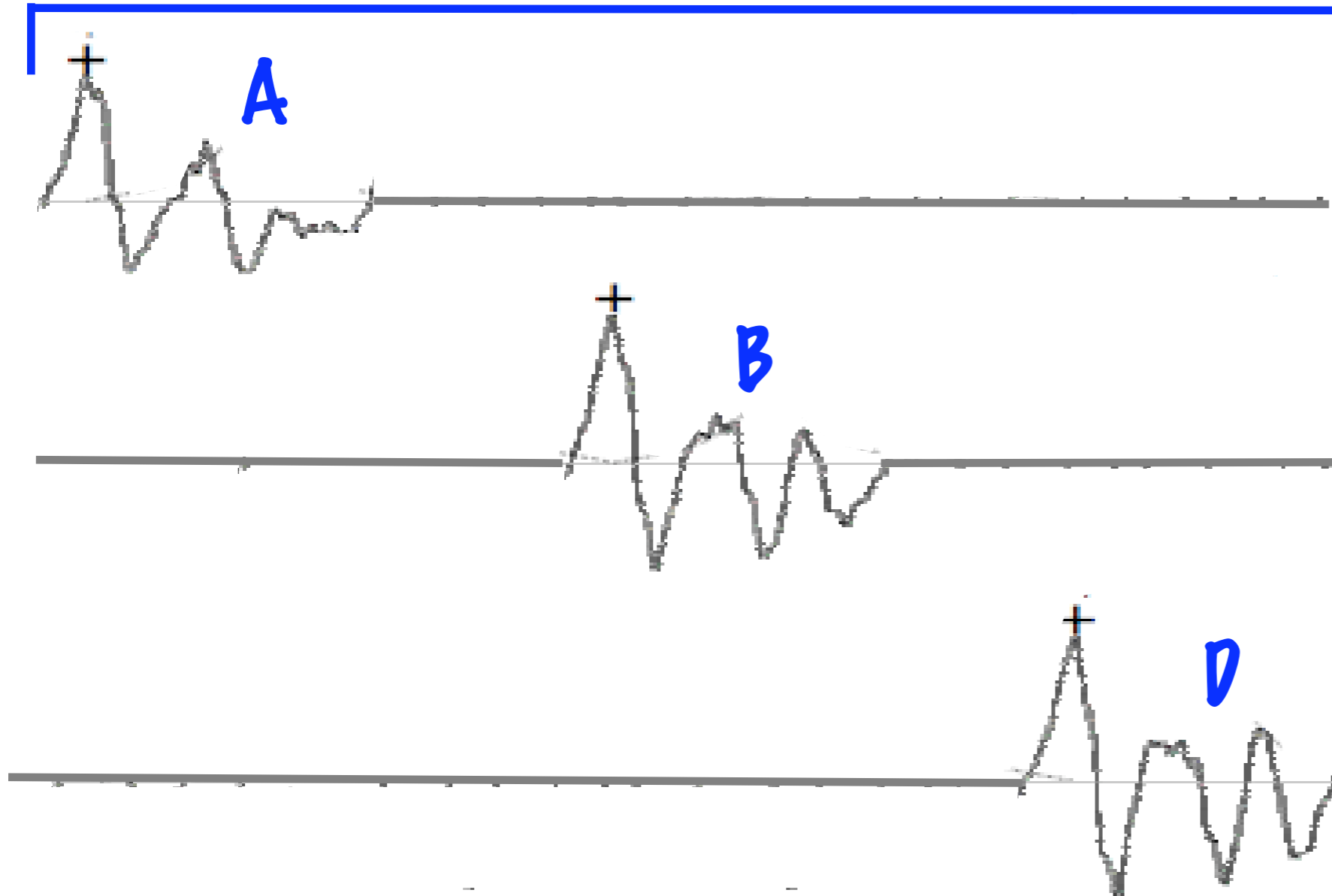
**Resampling preserves the magnitude of each partial.**

**Instead, we want formant frequencies to stay fixed ...**



# Recall: Beat-slicing pitch periods

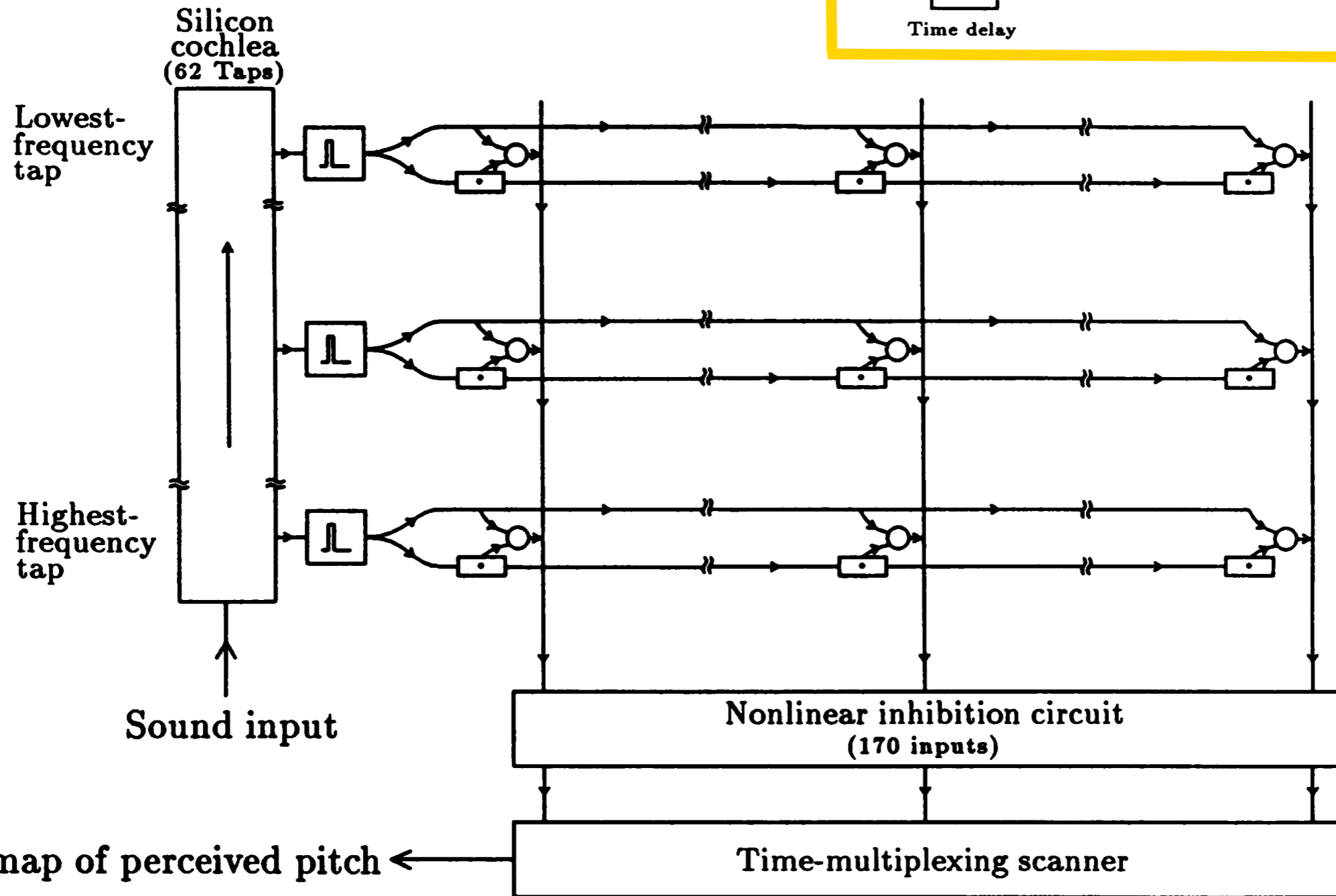
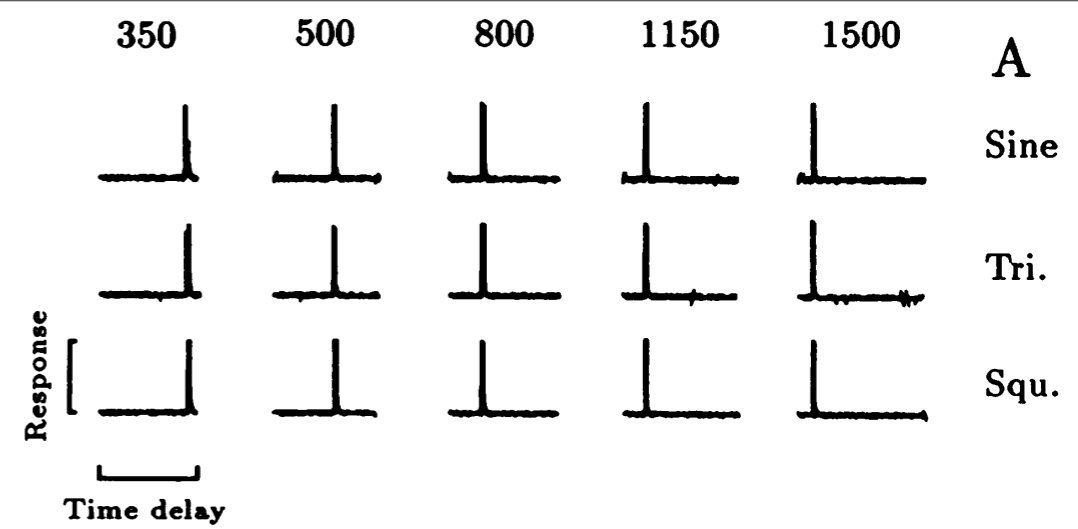
L



New local  
pitch:  
**Three**  
pitch  
periods  
per unit  
"L"

Apart from "edge artifacts", spectral shape  
is not changed by this operation ...

# Recall: Computing pitch



**Licklider model: Autocorrelate filtered waveforms.**



# Topics for today: Spectral processing

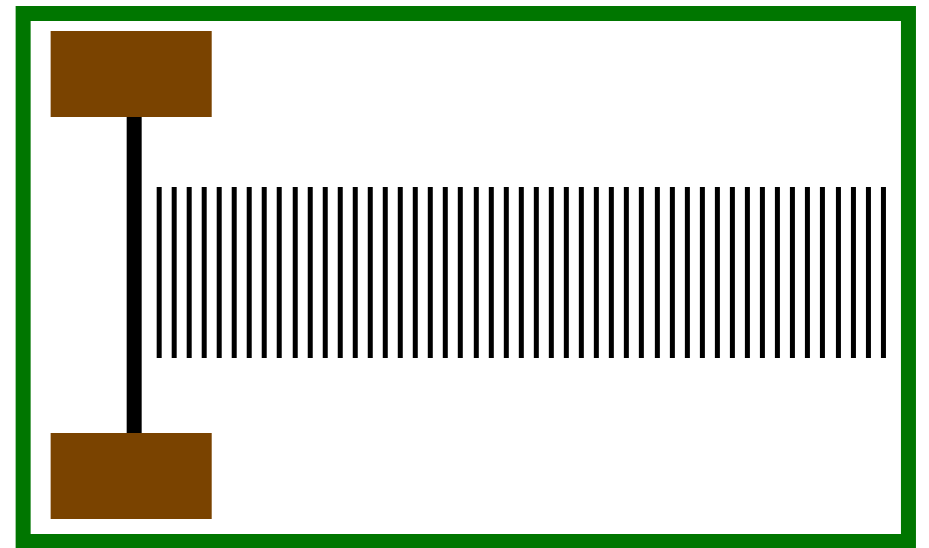
---

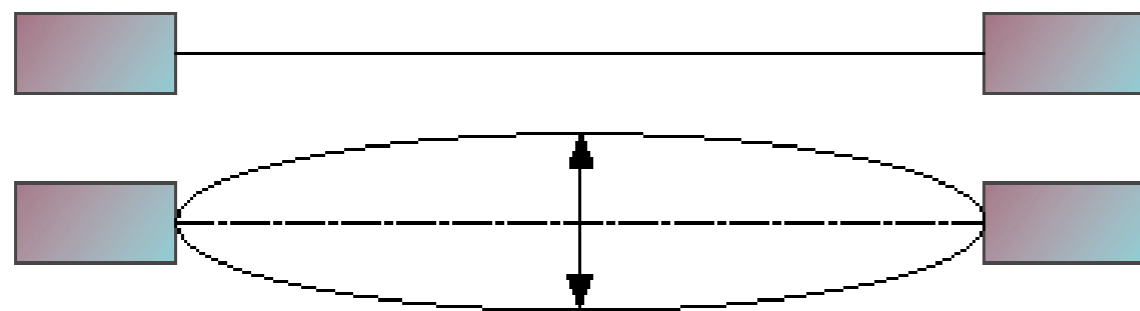
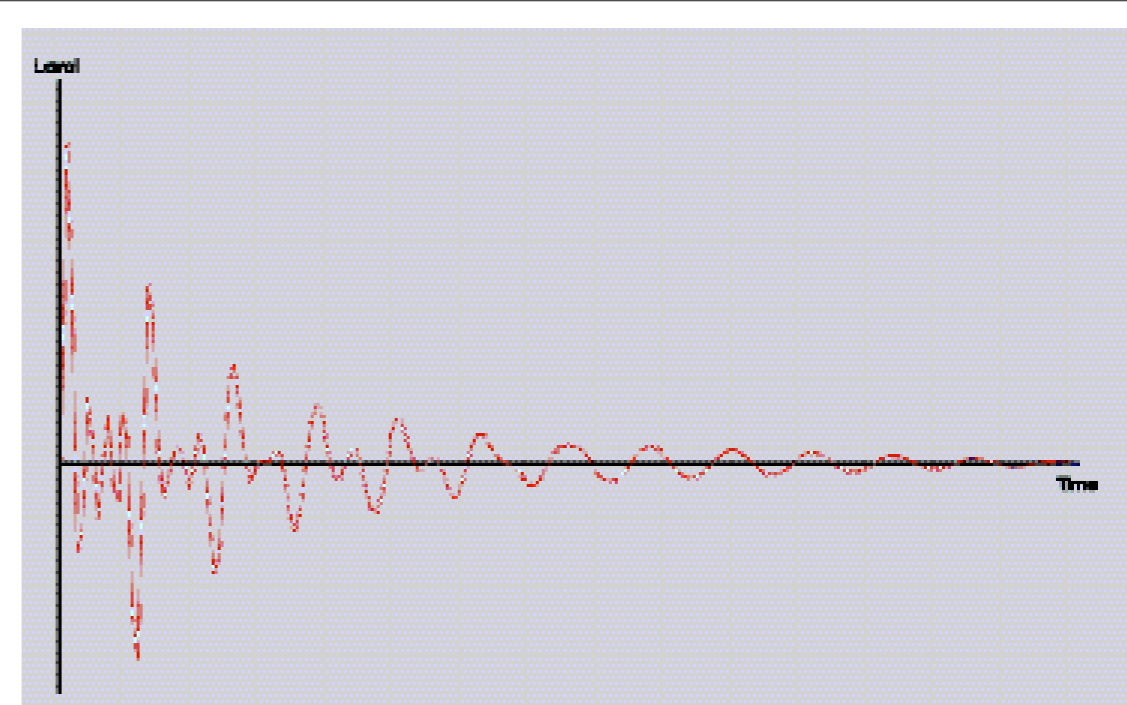
- \* **Analysis-synthesis:** Model sound as a set of parameterized sound generators.
- \* **Psychophysics:** Keeping sound object fusion as we modify the sound.
- \* **Time/Frequency Tradeoffs:** Narrow filters are slow, fast filters are wide.
- \* **Phase Vocoding:** The “Audio Image Processing” approach.

# Observation:

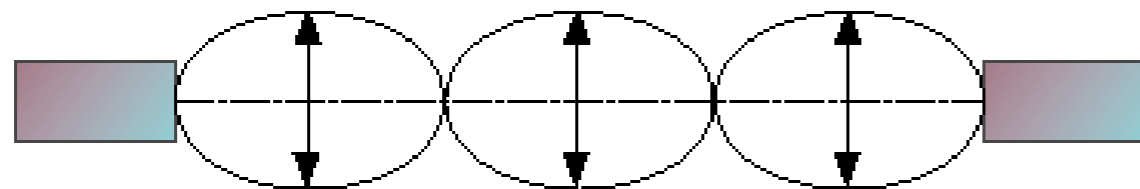
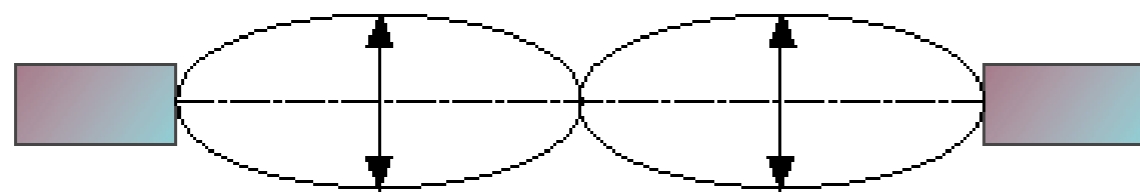
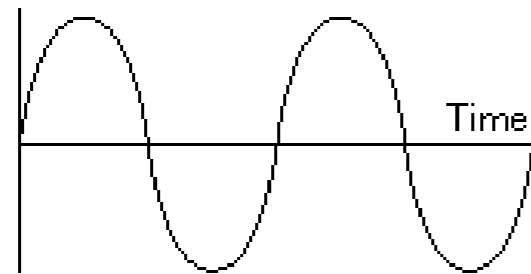
If you can fit a sound to a model with slowly-varying parameters,  
time-and pitch modification  
can be done in  
parameter space.

# Example: Plucked string

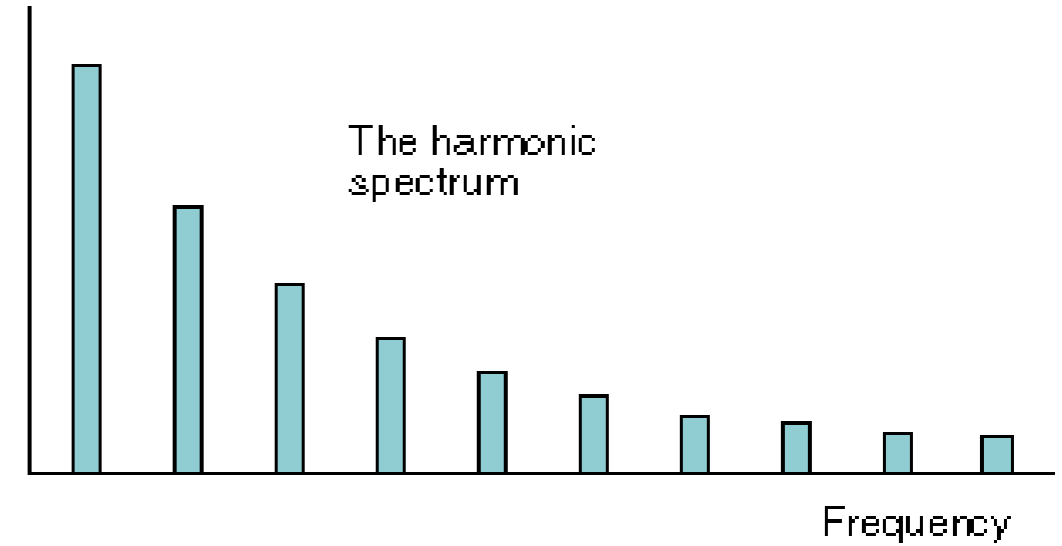




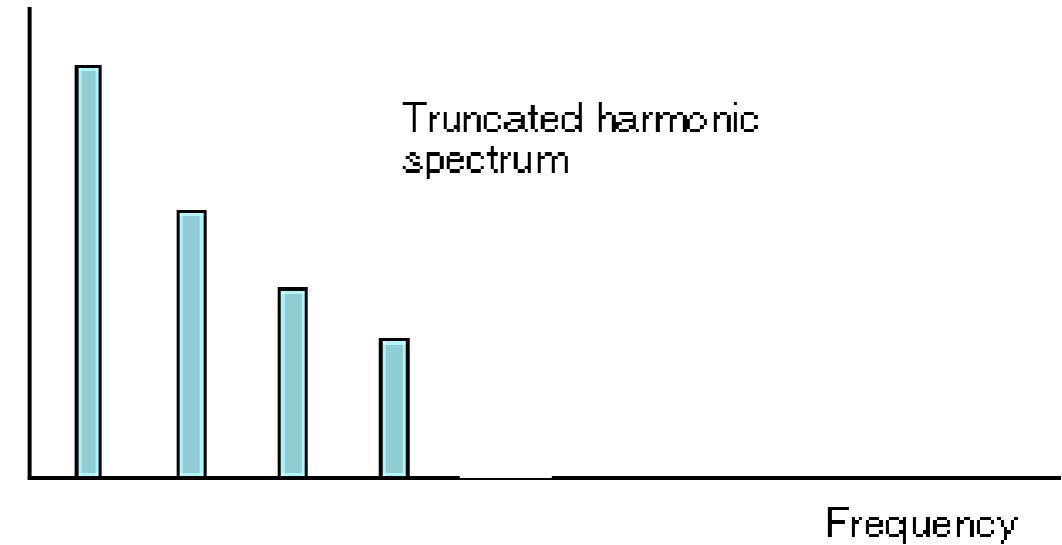
Displacement

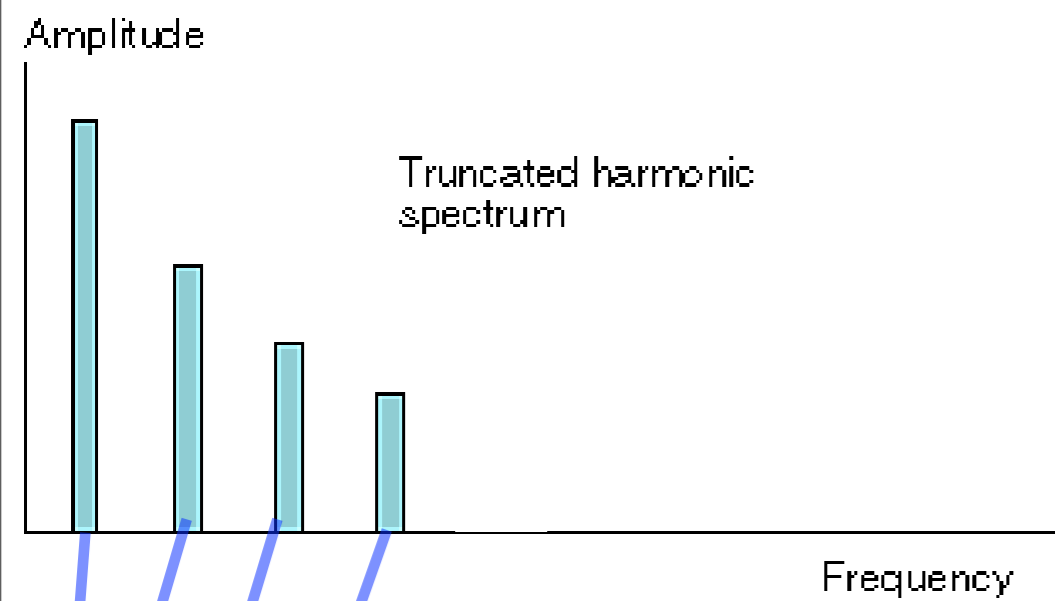


Amplitude



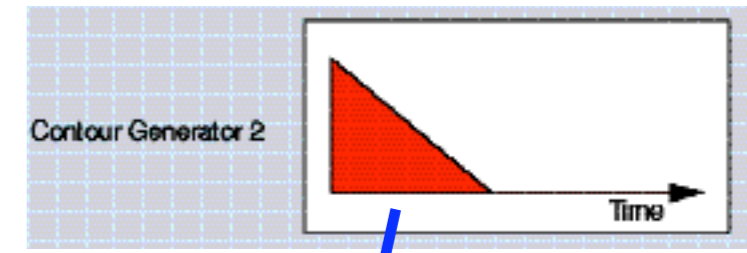
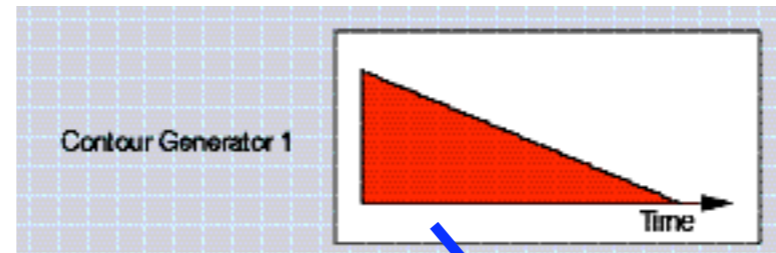
Amplitude





$A_1(t)$

$A_2(t)$

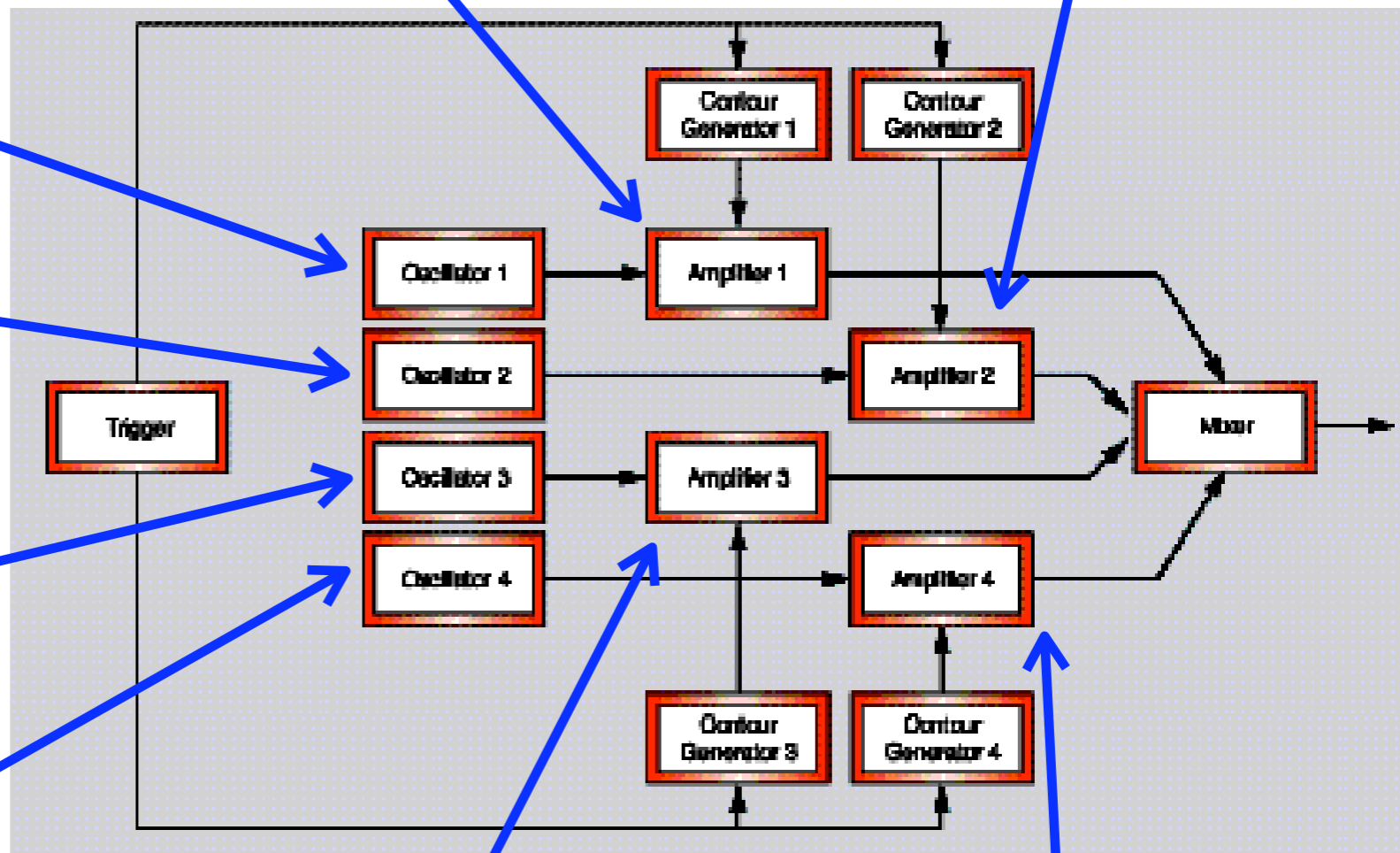


$$x(t) = A \sin(\omega t + \phi)$$

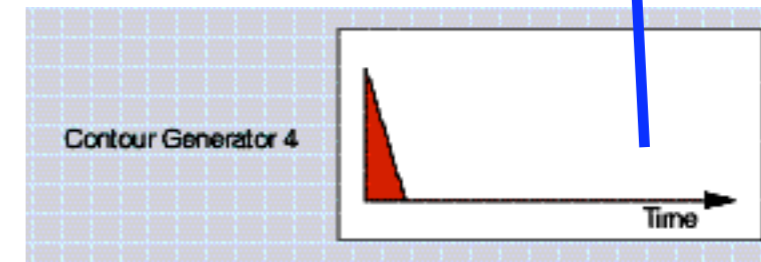
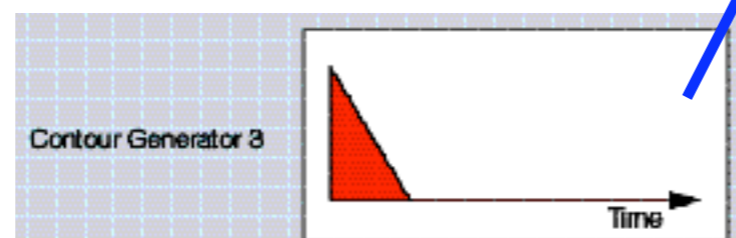
$$x(t) = A \sin(\omega t + \phi)$$

$$x(t) = A \sin(\omega t + \phi)$$

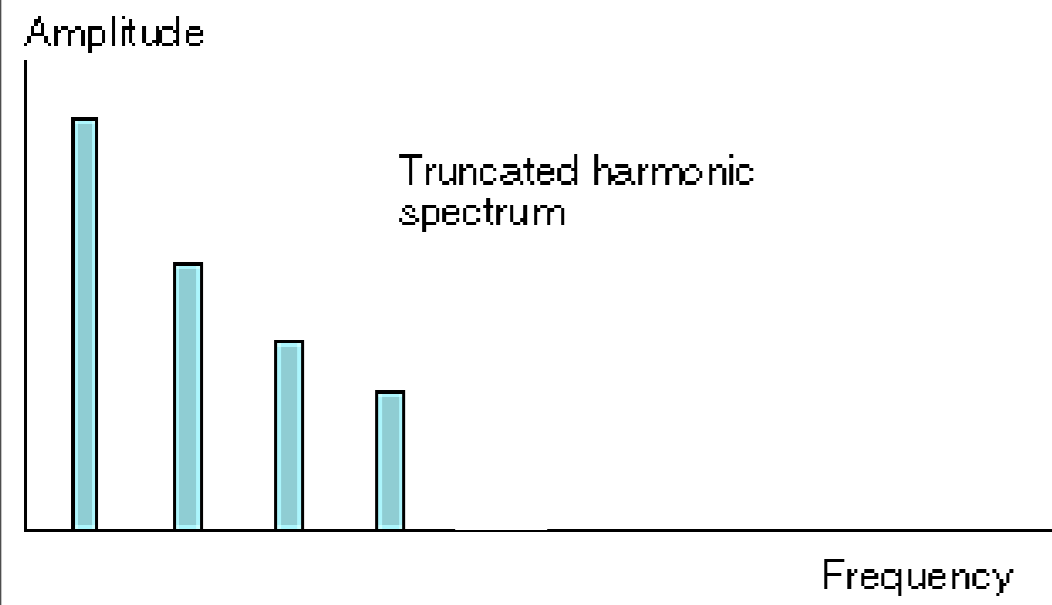
$$x(t) = A \sin(\omega t + \phi)$$



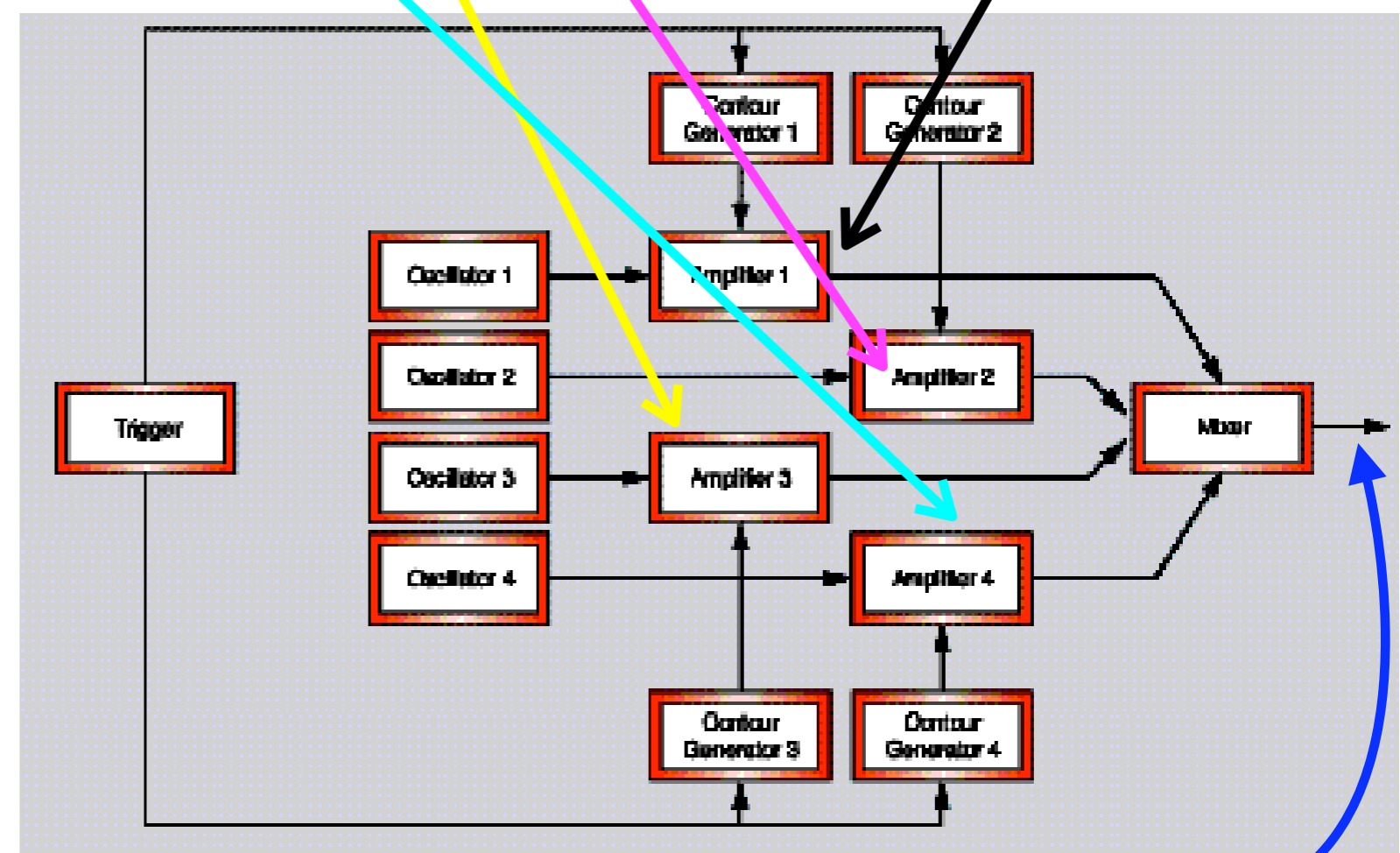
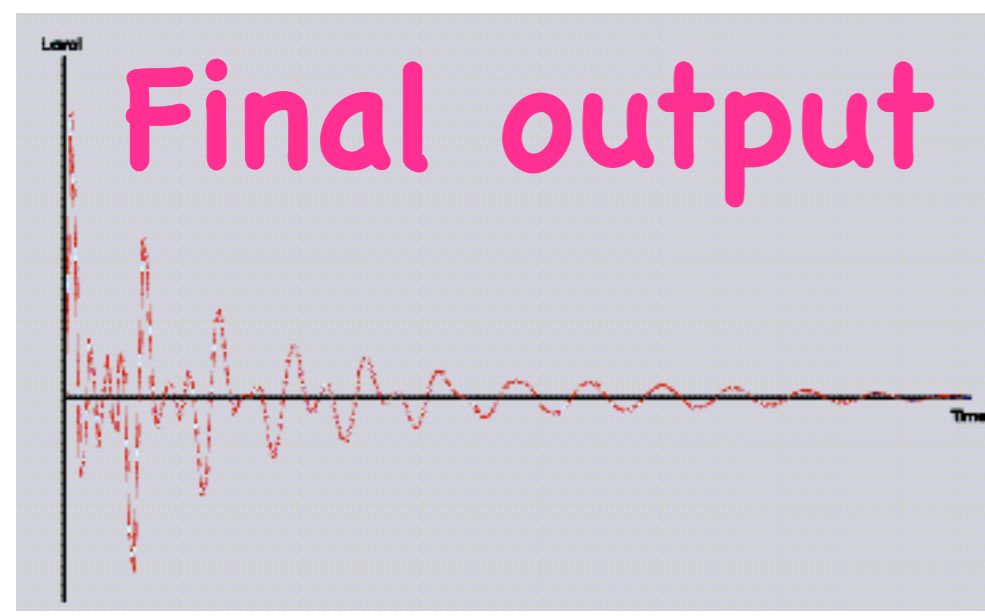
$A_3(t)$

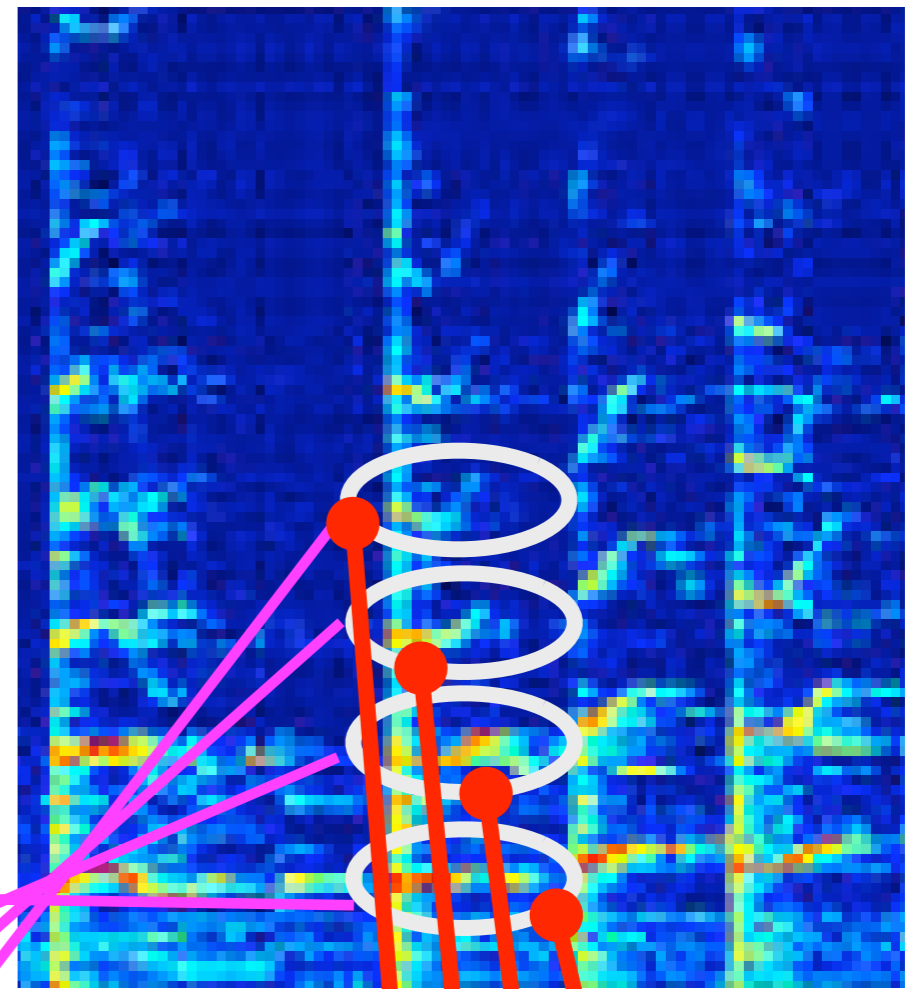


$A_4(t)$



"Additive synthesis"



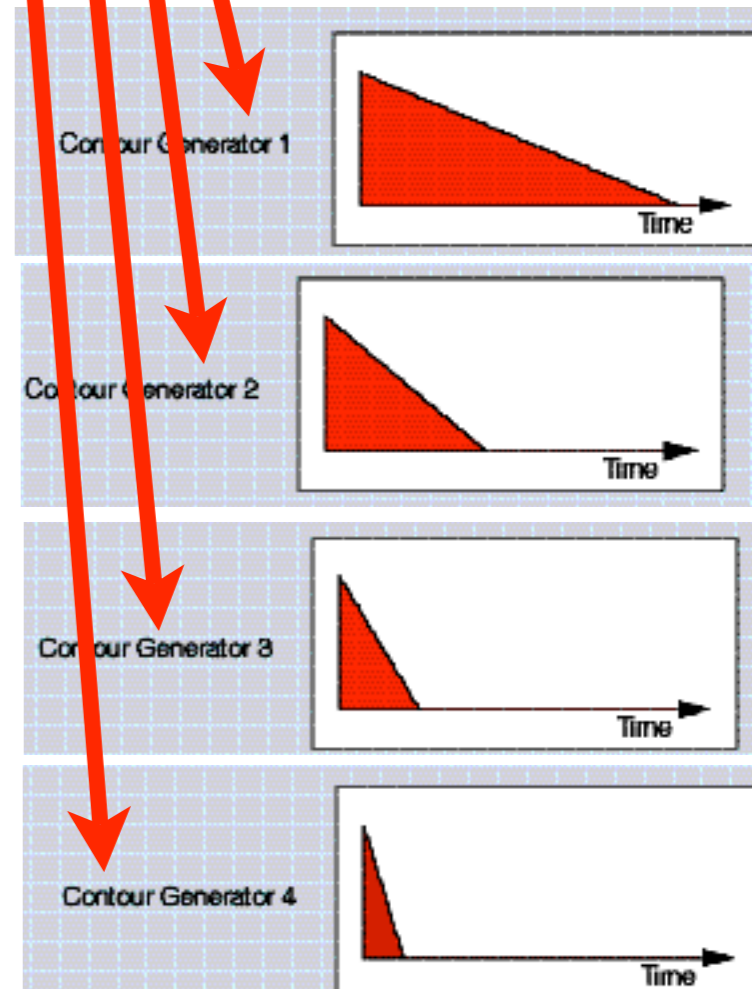


$$x(t) = A \sin(\omega t + \phi)$$

$$x(t) = A \sin(\omega t + \phi)$$

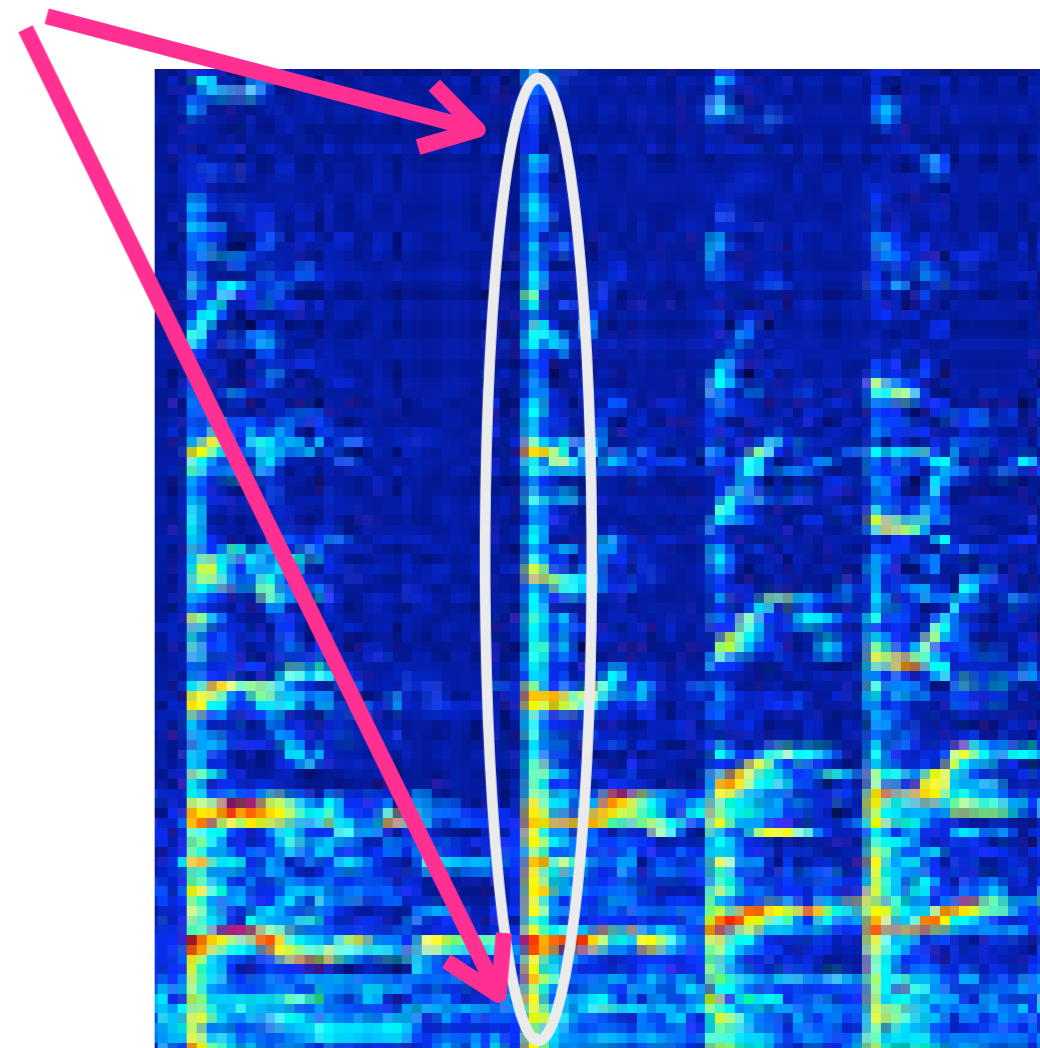
$$x(t) = A \sin(\omega t + \phi)$$

$$x(t) = A \sin(\omega t + \phi)$$



**Manage lifecycle of tracks:  
Birth, evolution, death.**

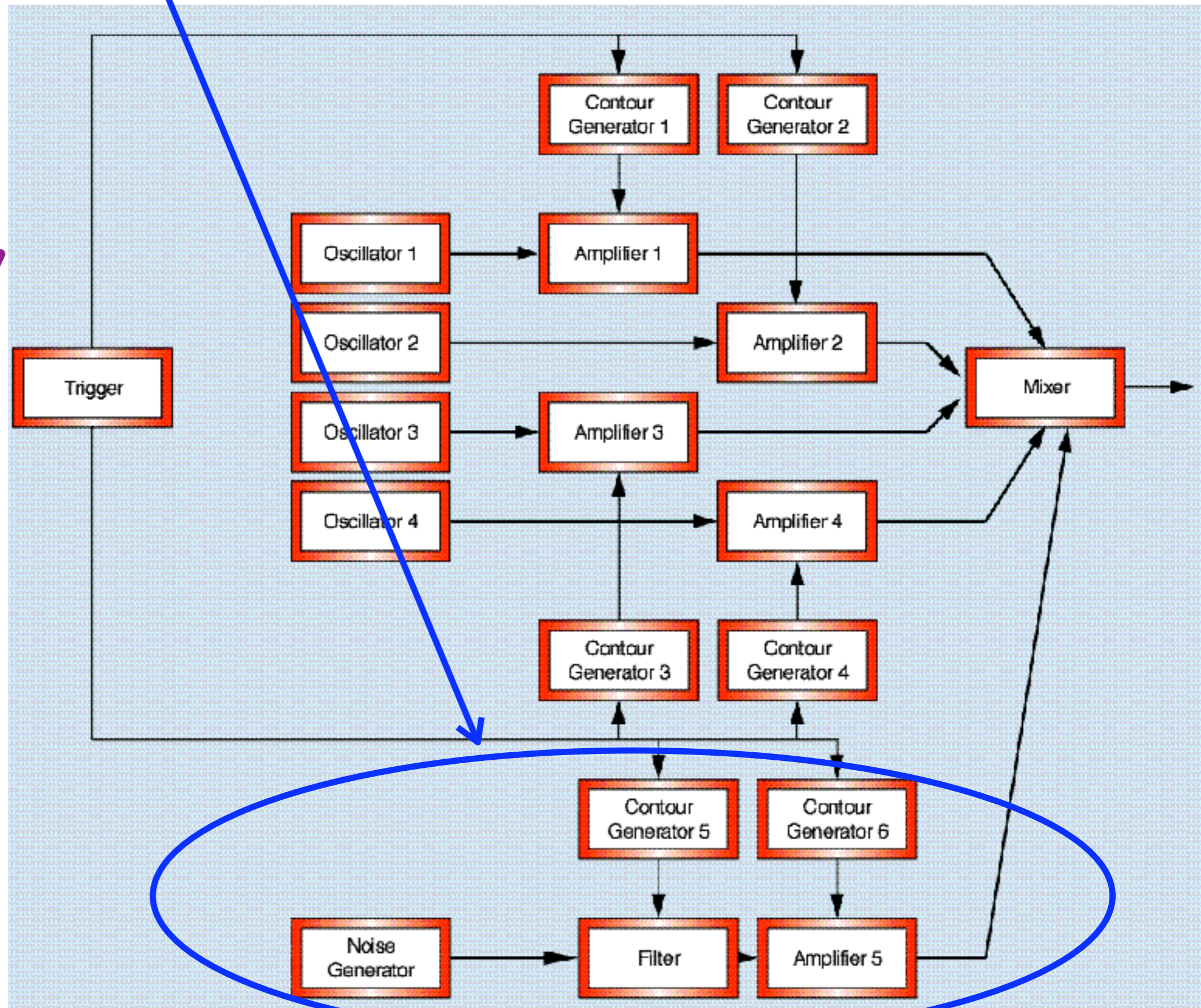
Some sound components are not well modelled by sinusoids.





# Sine + filtered noise models ...

Sine +  
transients,  
Sine +  
residue,  
etc ...



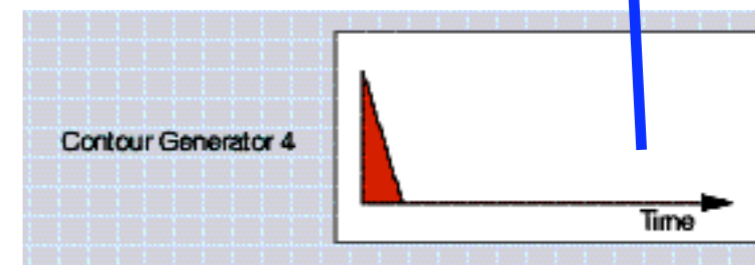
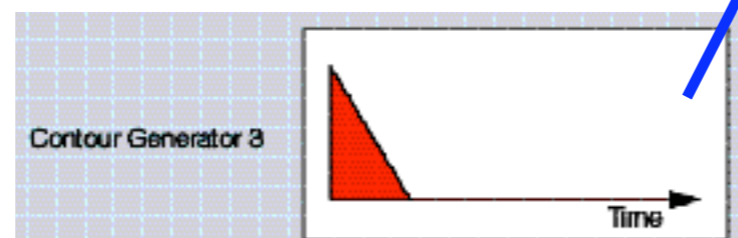
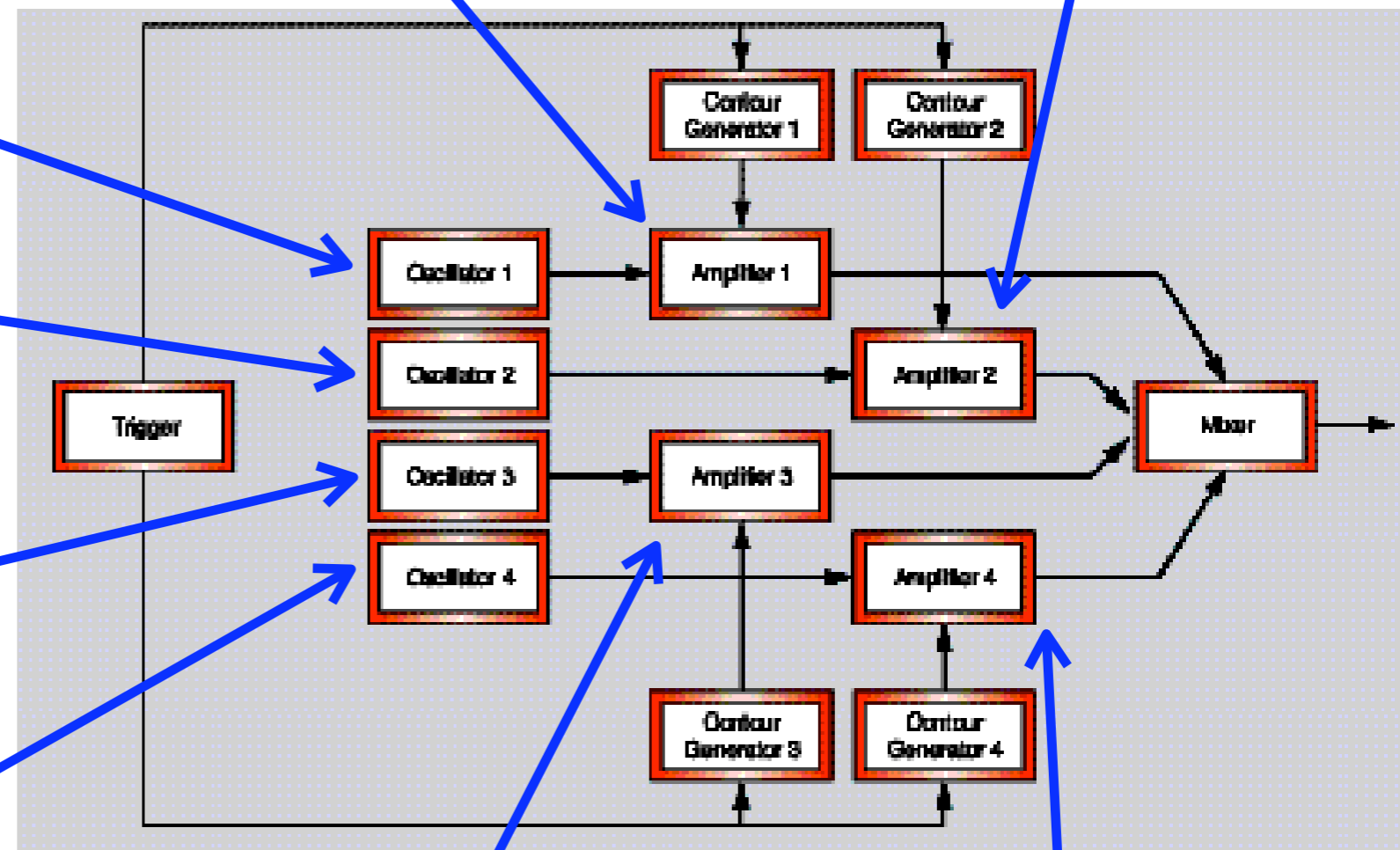
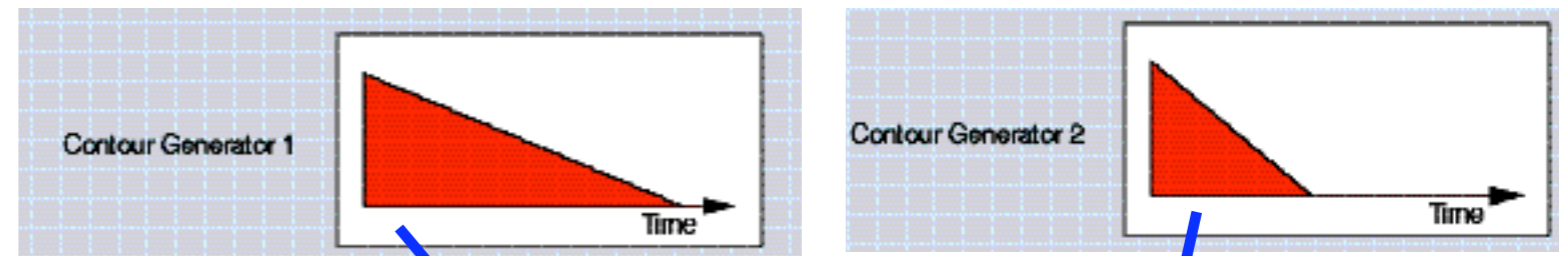
Recall: If you can fit a sound to a model, time-and-pitch modification can be done in parameter space.

$$x(t) = A \sin(\omega t + \phi)$$

$$x(t) = A \sin(\omega t + \phi)$$

$$x(t) = A \sin(\omega t + \phi)$$

$$x(t) = A \sin(\omega t + \phi)$$



# Commercial example:

## Celemony Melodyne resynthesis editor.

The screenshot displays the 'ueberschall' software interface by Liquid Instruments. The top left features the 'Liquid Instruments' logo. Below it, a text box contains 'sax/tenor/pop/D#' and a tempo control set to '120.00'. The top right section includes a 'ueberschall' title and four sliders for 'volume', 'panorama', 'pitch', and 'formant', each with a corresponding dropdown menu (solo, file, action, setup). Below these are 'sounds' and 'editor' tabs. The main interface includes a 'play' button, a 'cycle' checkbox, a 'grid' set to '1/4', a 'snap' set to 'semi', and a key signature of 'D# major'. The central piano roll editor shows a green grid with a vertical axis for notes (C, C#, D, D#, E, F, F#, G, G#, A, A#) and a horizontal axis for time divided into four measures. A tempo of 120.0 is indicated at the start. Several notes are visible as horizontal bars with green fill and white outlines, indicating their duration and pitch across the measures.

powered by melodyne

# Commercial example: Synful Orchestra

“about 100 sinusoids  
per note, + noise models”

The screenshot displays the Synful Orchestra software interface. At the top left, there is a grid of 16 instrument selection buttons, numbered 1 through 16, with dropdown menus for each. The instruments listed are: 1. violin, 2. violin, 3. viola, 4. cello, 5. flute, 6. oboe, 7. english horn, 8. clarinet, 9. clarinet, 10. trombone, 11. bassoon, 12. horn, 13. horn, 14. trumpet, 15. trumpet, 16. trombone.

To the right of the instrument grid, the word "SYNFLUL" is written in large yellow letters, with "ORCHESTRA" in smaller black letters below it. Below the title, there are two buttons: "single channel view" and "help".

Below the instrument grid, there is a "tune orchestra" section with a numerical display showing "440.00" and a slider. To the right of this is a "delay for expression" button with a red "on" indicator.

The "Program Parameters" section contains several controls:

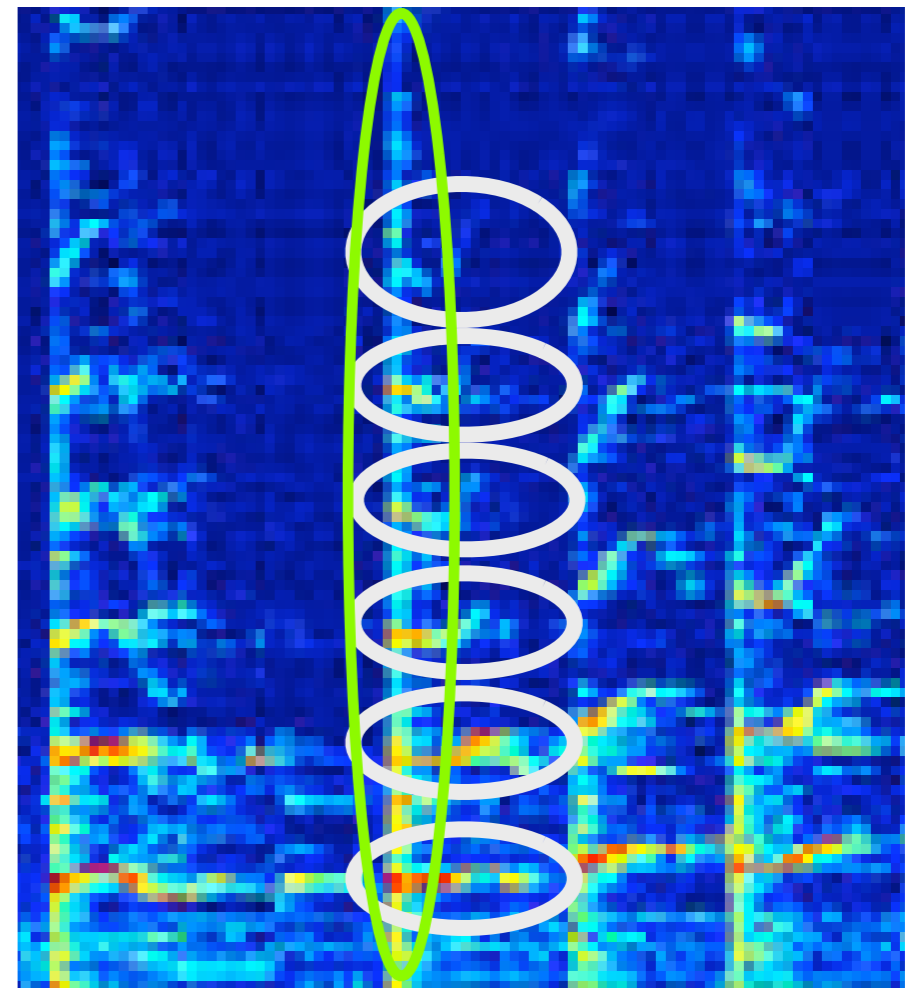
- "copy program" and "delete program" buttons.
- "Synful pitch wheel" button with a red "on" indicator.
- "pitch wheel range" slider with a value of "12.00".
- "gain" slider with a value of "-1.58".
- "harmonic BT" slider with a value of "-15.44".
- "transient gain" slider with a value of "4.11".
- "release trim" slider with a value of "-0.37".
- "harmonic parity" slider with a value of "-0.88".
- "sustain noise trim" slider with a value of "-2.68".
- "tune instrument" slider with a value of "0.00".

At the bottom left, it says "Copyright 2004-2005 Synful LLC". At the bottom right, it says "SYNFLUL ORCHESTRA VERSION 2.1.1".

# Concept: **Common Fate**

---

Changes to sine tracks and other components that belong to the same note should not "break" object percept.

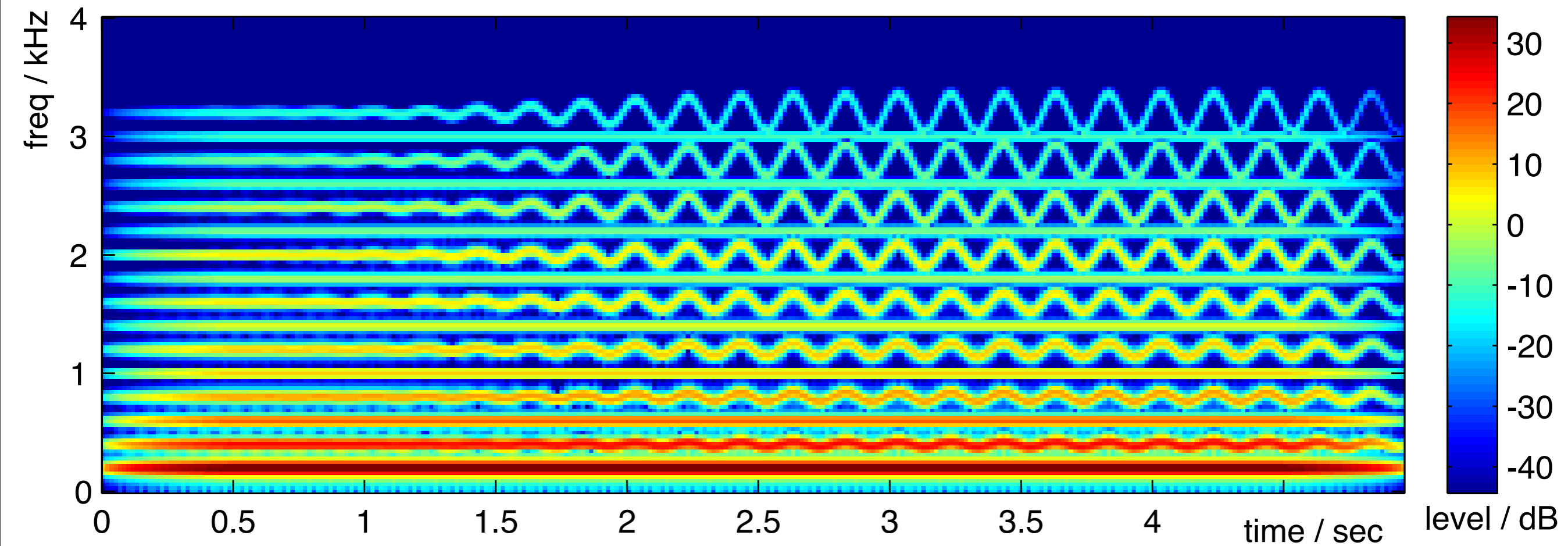


Pitch changes (vibrato, bends).

Phase continuity (noise to sines).

Amplitude envelope relations.

# Example: Reynolds/McAdams Oboe



**Adding vibrato to even partials** **As shown**  
**makes them separate from the**  
**odd partials. (spectrum by Dan** **Reduced**  
**Ellis).**

# Recall: **Fusing onset snippets**

---

## SYNTHESIZING TRUMPET PERFORMANCES

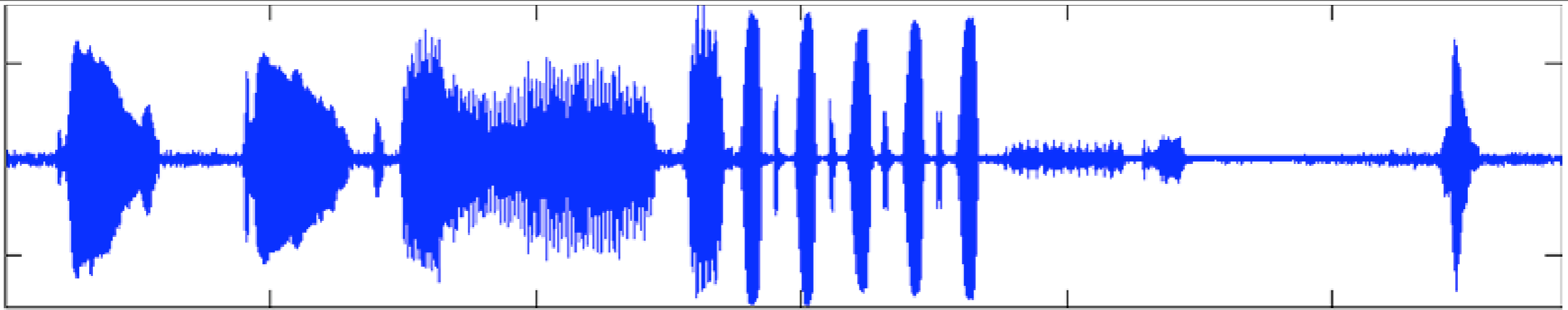
Istvan Derenyi and Roger B. Dannenberg

School of Computer Science, Carnegie Mellon University

Pittsburgh, PA 15213, USA

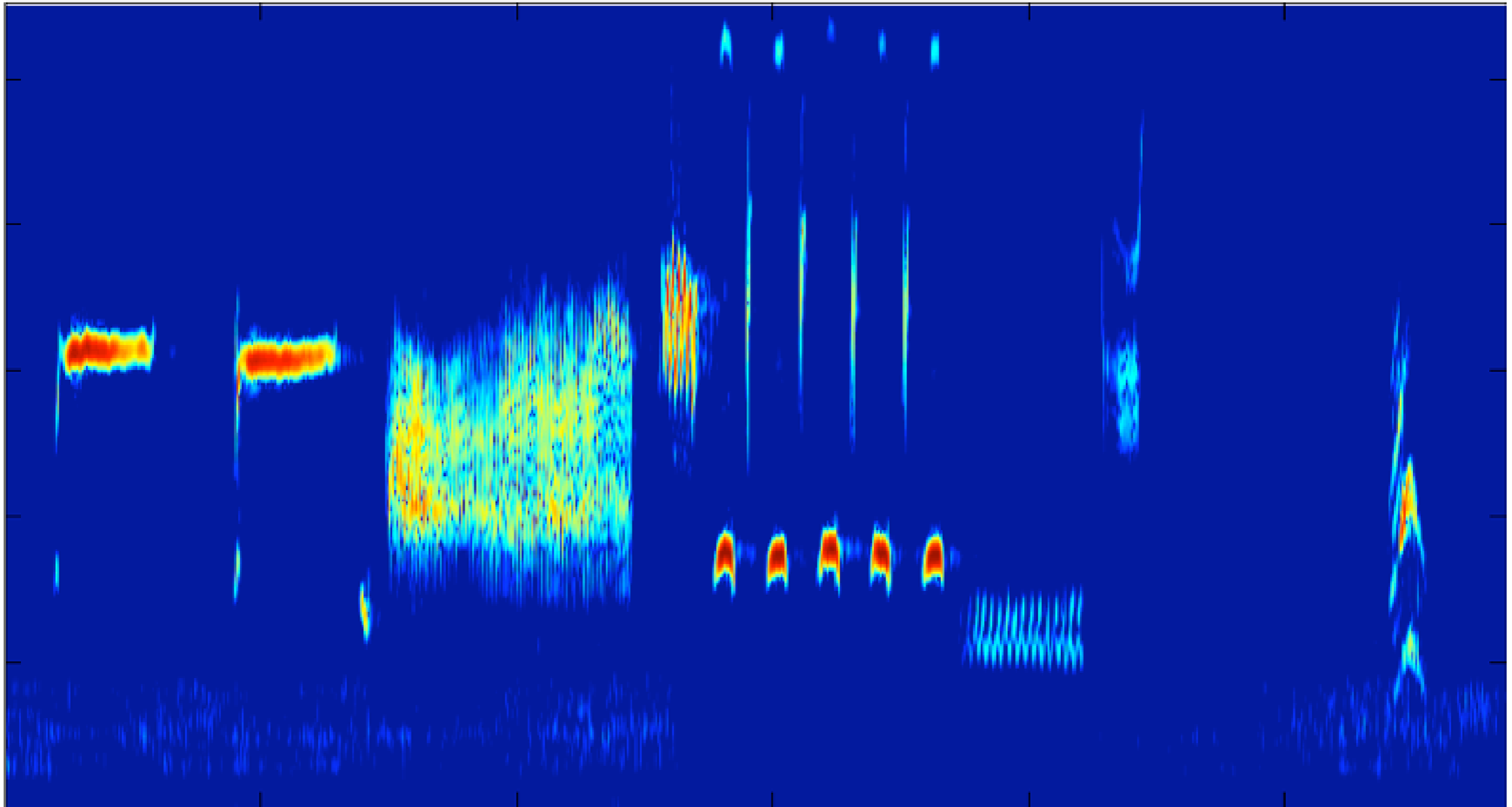
{derenyi, rbd}@cs.cmu.edu

- \* Each scale note has a trumpet onset sample.
- \* **Measure** the amplitude and phases of trumpet harmonics at the **end of onset sample**.
- \* To begin the sustained sound, a waveform is **calculated** whose phases and amplitudes **match the onset**.
- \* Over 50 ms, **interpolate** to the **desired** amplitude **spectrum** of the sustained sound.



Sparrow Bird Call

Play



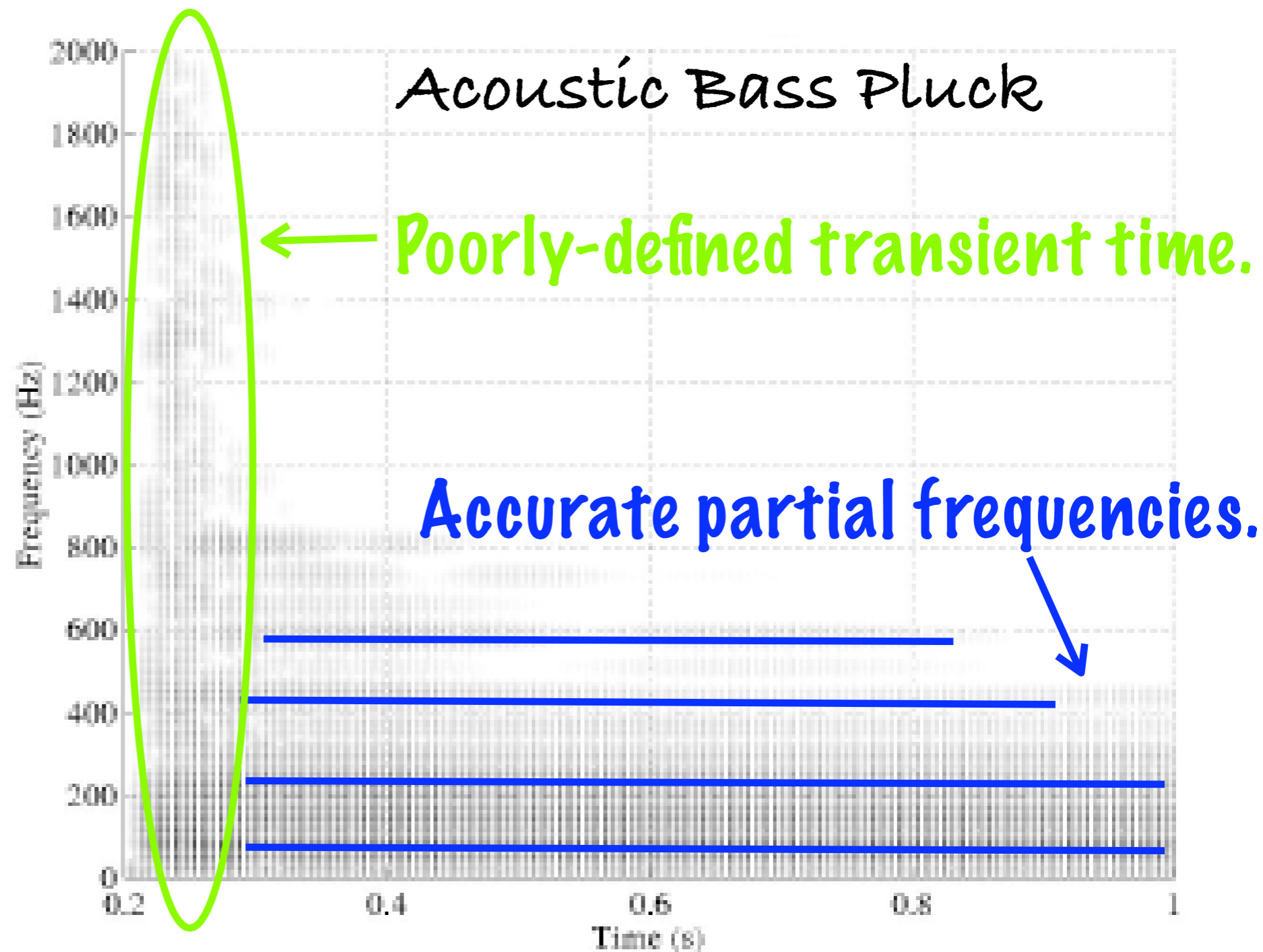


# Time/Frequency Tradeoffs

Fast linear filters are wide.

Narrow linear filters are slow.

Playing

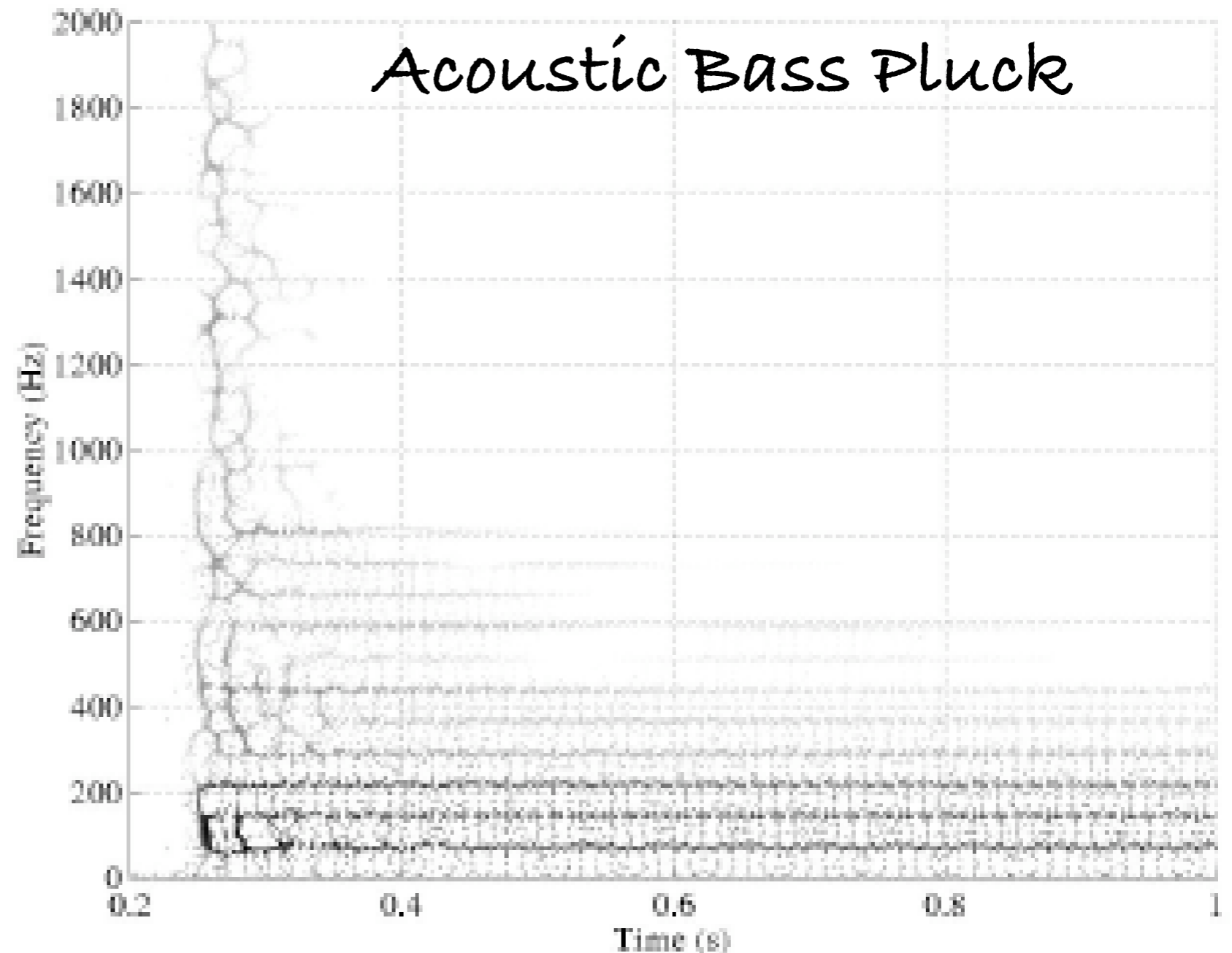


# Time/Frequency Reassignment

Dedicate phase of spectrum to adjust center of mass of each "grid point".

Not a free lunch.

No phase information for the reassigned spectrum.



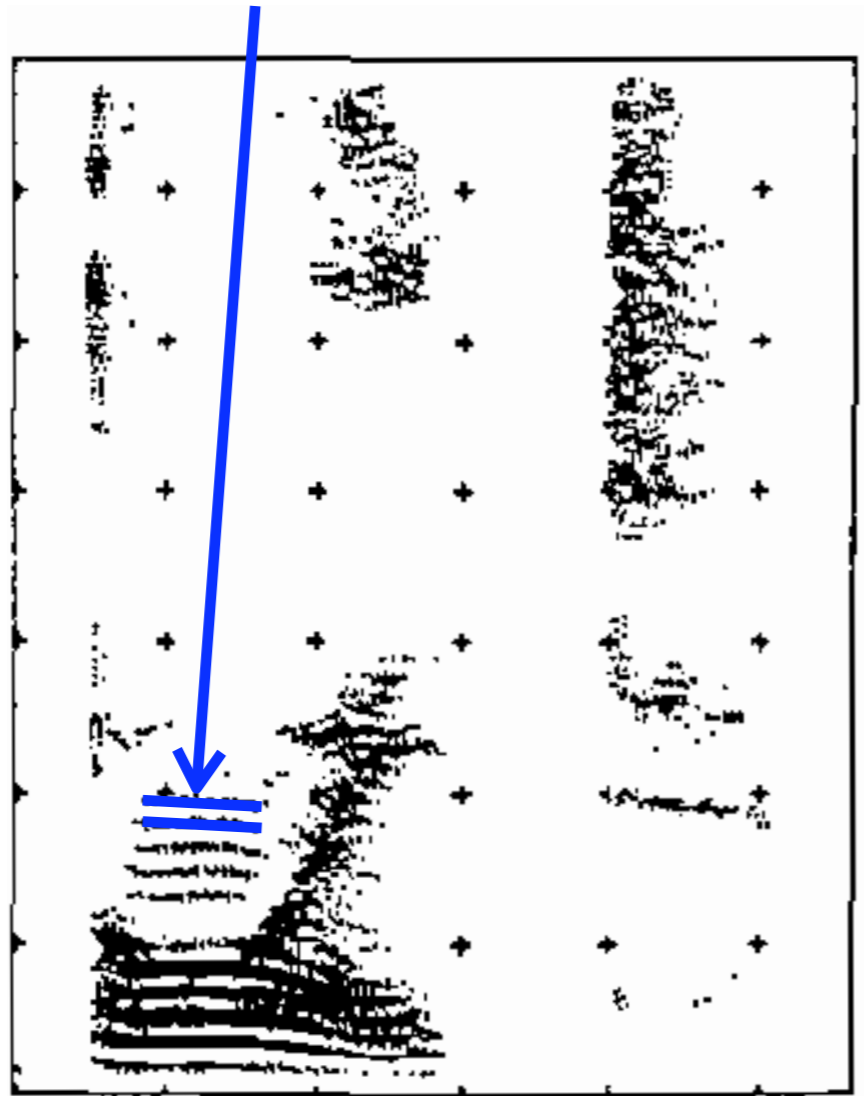
Data from Kelley Fitz and Sean Fulop

# Scale-Space Approach

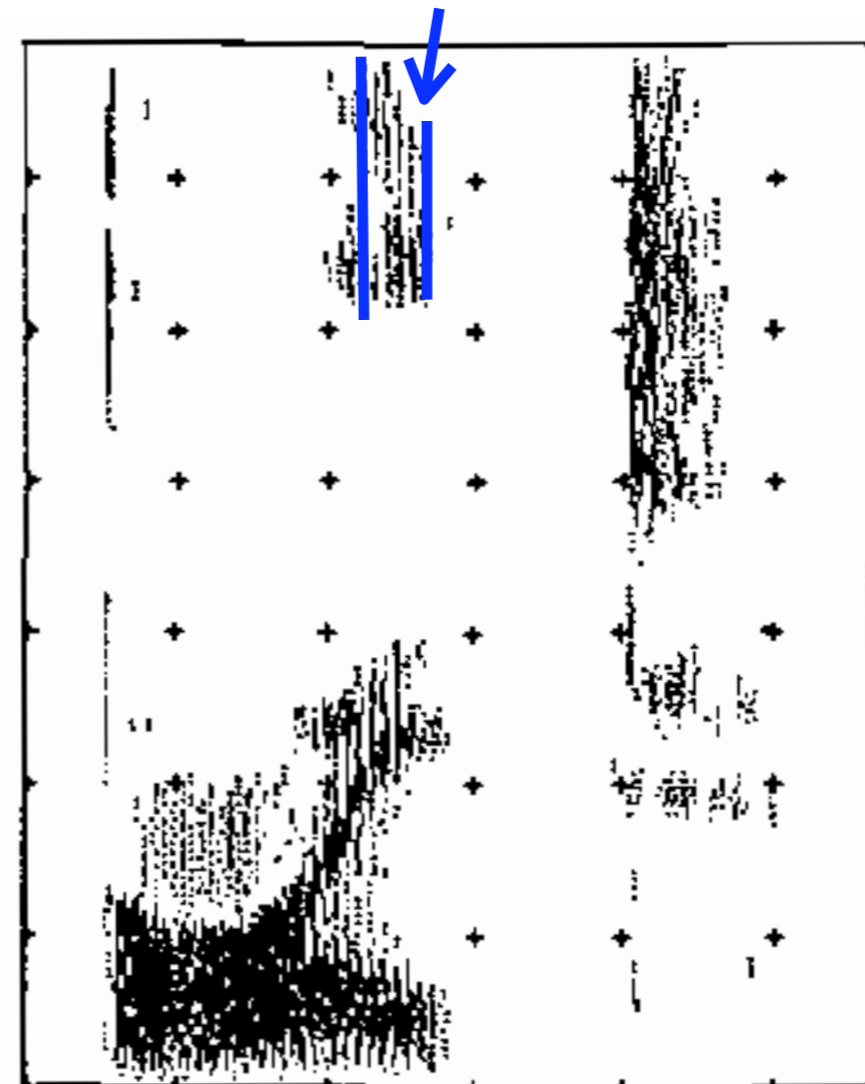
Spoken word "boyt"  
processed by **two**  
filterbanks.

Accurate pitch harmonics.

Good glottal pulse timing.



Narrow filters

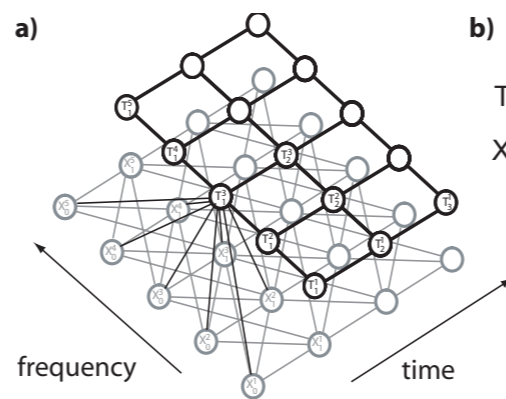


Fast filters

Data from Hong Leung and Victor Zue

# Multi-Level Maps

Data from Reyes-Gomez, Jojic, and Ellis



Derive motion maps from spectrogram.

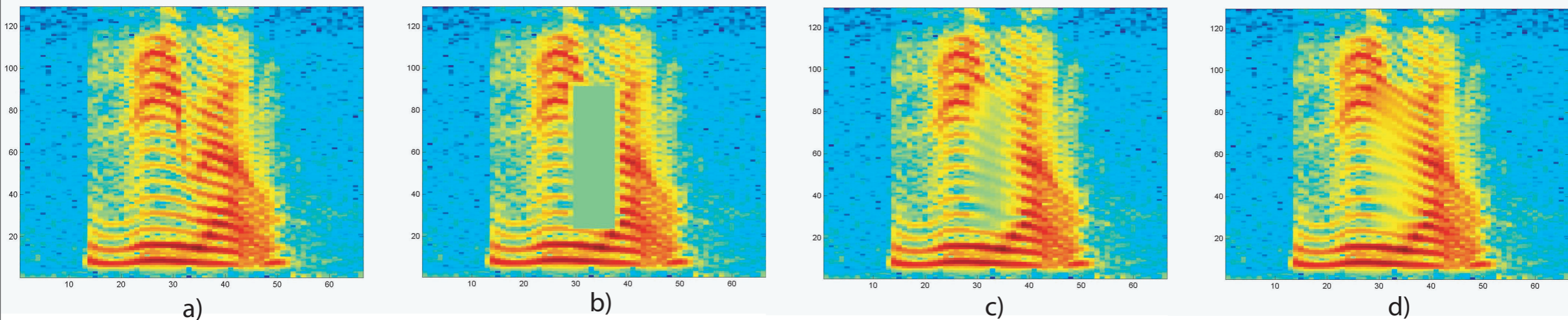
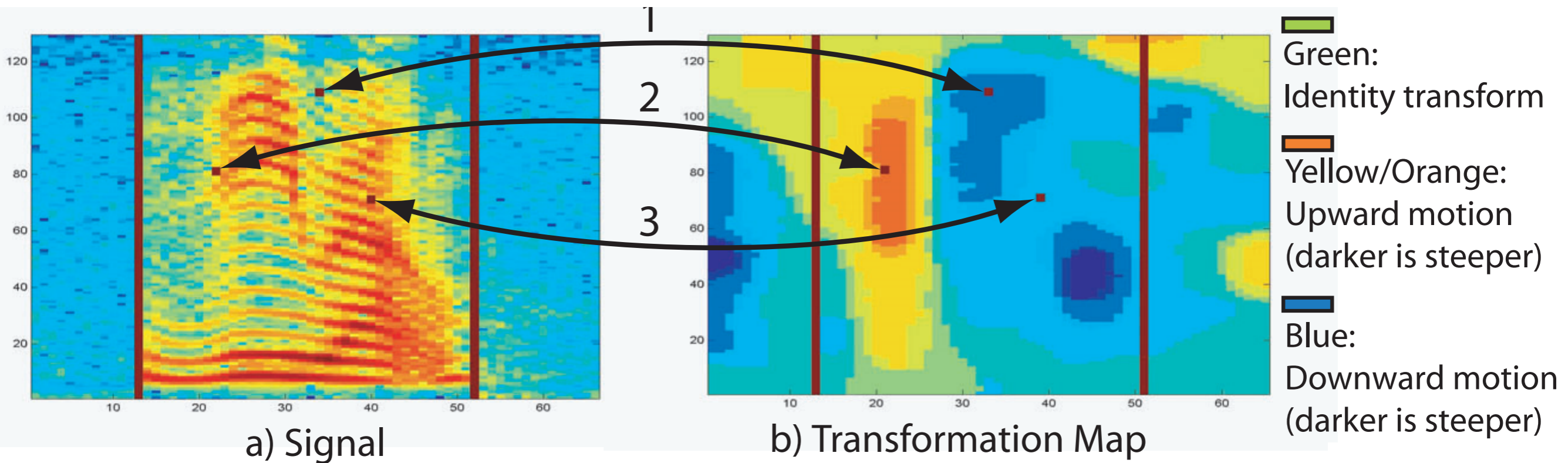


Figure 3: Missing data interpolation example a) Original, b) Incomplete, c) After 10 iterations, d) After 30.

# My Former Life ...

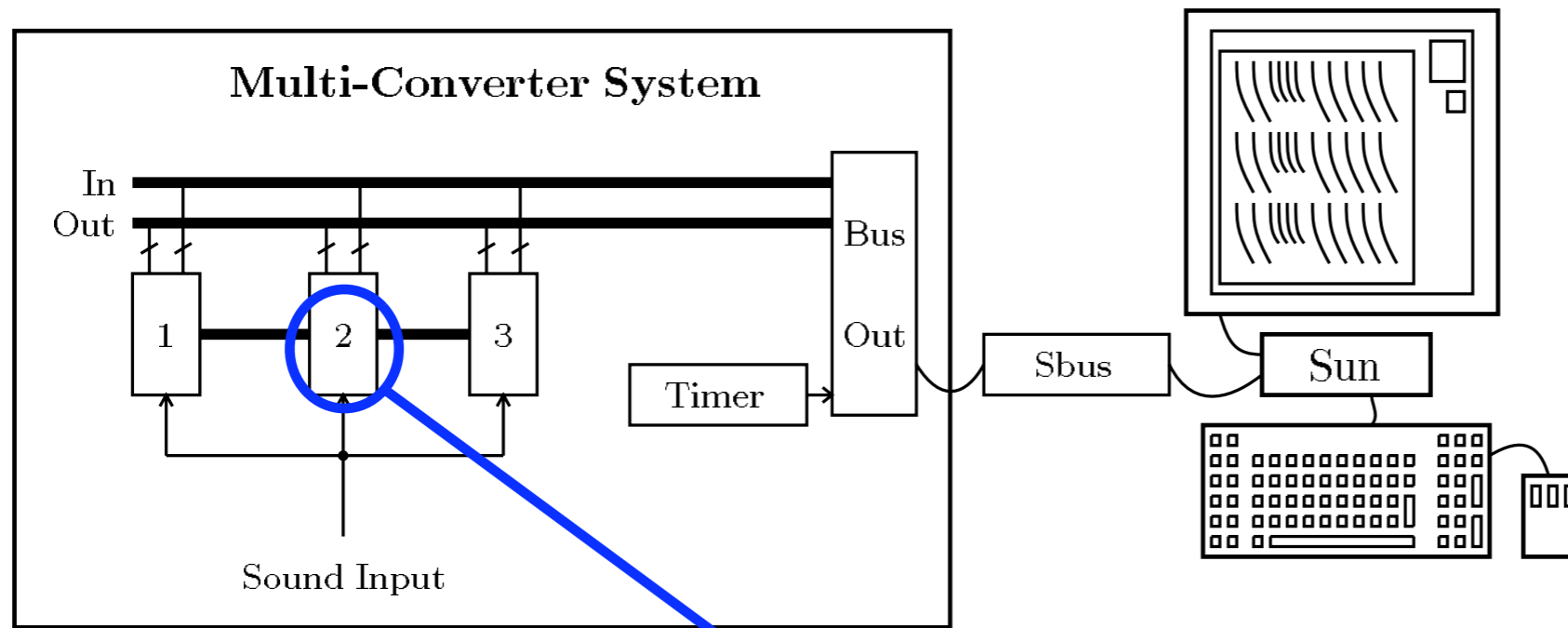


Figure 1. Block diagram of the multi-converter system.

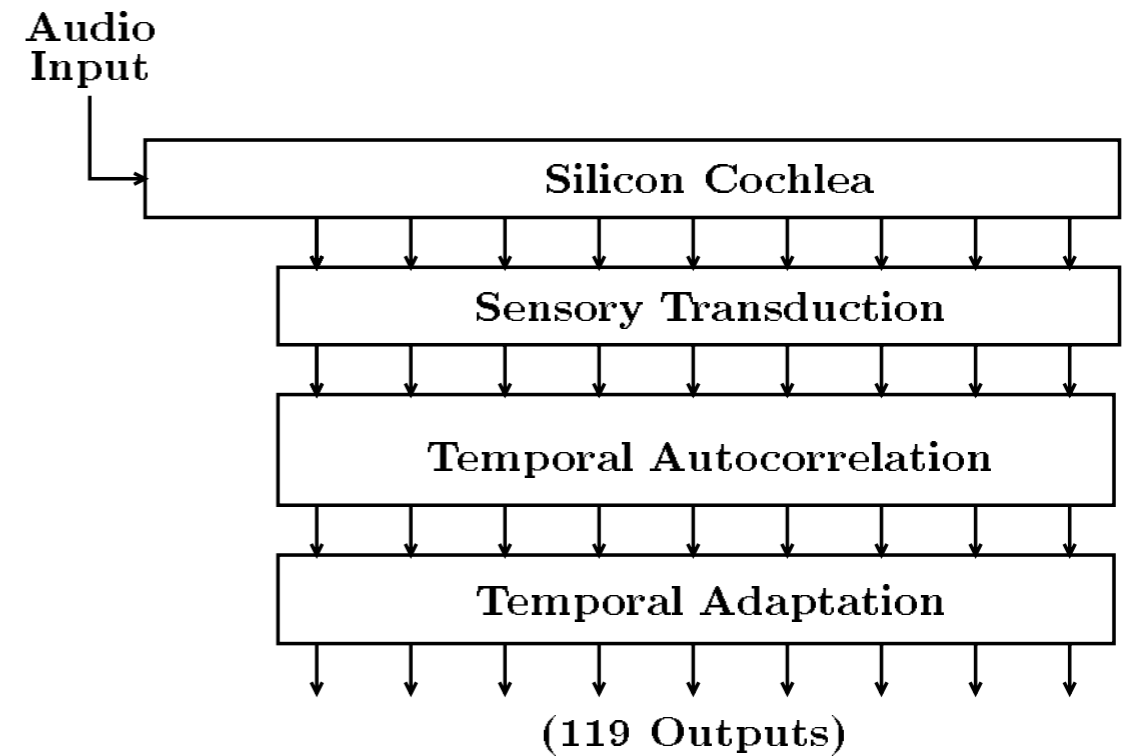


Figure 2. Analog signal path of the silicon auditory model.

# My Former Life ...

## Auditory Models

Adaptive Sampling

Specialized Features

Multiple Representations

High-Dimensional

Correlated Features

## Speech Recognition

Uniform Sampling

General-Purpose Features

Single Representation

Low-Dimensional

Uncorrelated Features

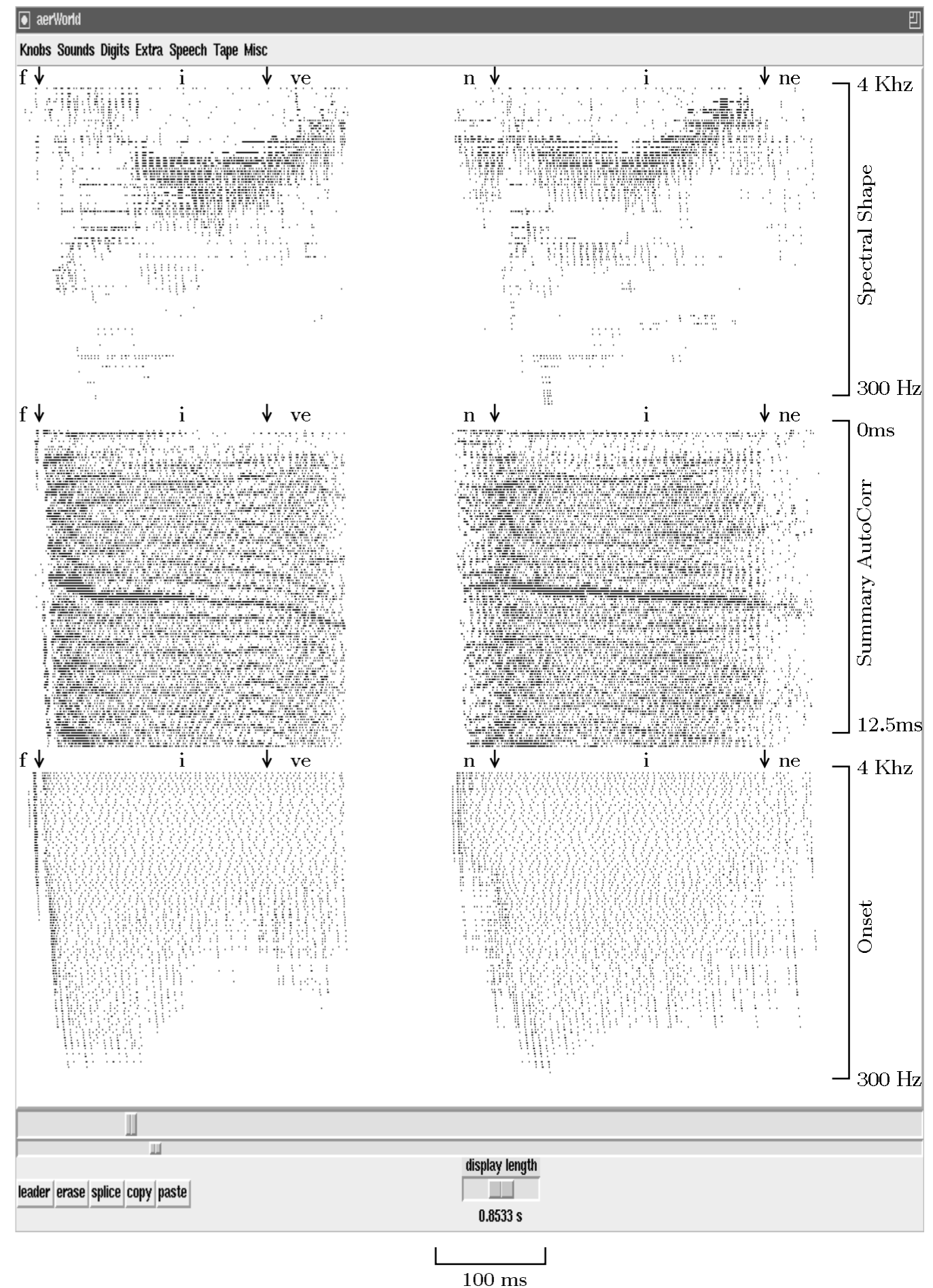


Figure 7. Data from the multi-converter system, in response to the word “five” followed by the word “nine”.

# My Former Life ...

Features	Parameters	Hidden Units	1	2	3	4	Average
SS	65,586	326	6.6	6.9	5.4	8.0	6.7
SS + Auto	65,468	276	5.7	5.8	4.5	5.5	5.4
SS + Onset	65,531	225	4.9	5.1	4.3	4.9	4.8
SS + Auto + Onset	65,456	200	4.9	4.2	3.2	4.0	4.1

**Figure 12.** Percent error for feature vectors derived from auditory representations (four database partitions). Other fields show number of hidden units and number of parameters in the MLP classifier net. Code: SS = spectral shape features, Onset = onset features, Auto = autocorrelogram features.

Features	Total	9/5	oh/no	others
SS	6.7	1.4	1.0	4.4
SS + Auto	5.4	1.1	0.8	3.5
SS + Onset	4.8	1.0	0.7	3.1
SS + Auto + Onset	4.1	0.7	0.6	2.8

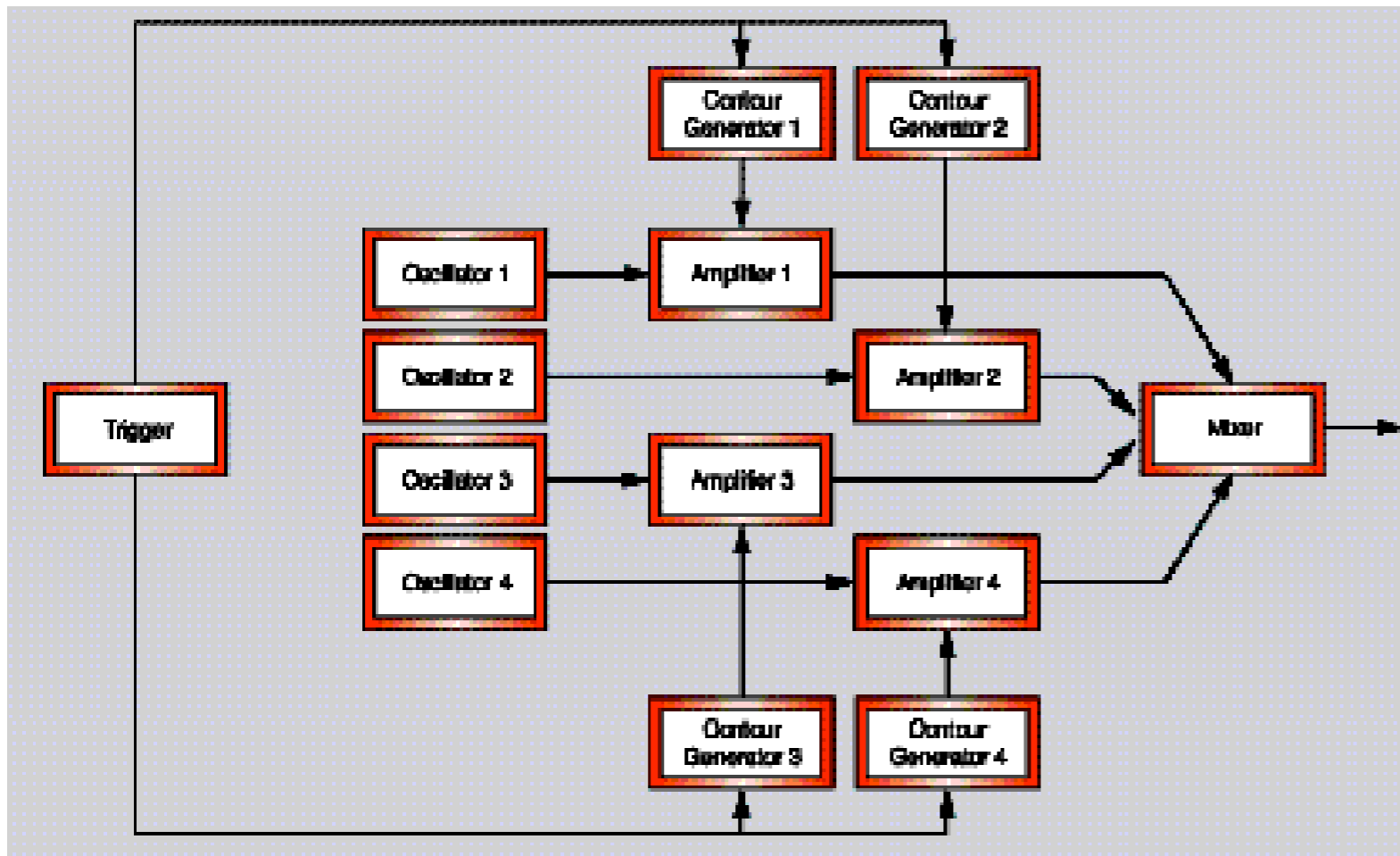
**Figure 13.** Error analysis of the recognition experiments in Figure 12 (averaged over partitions). Errors due to the two leading word confusions are listed (confusing “five” and “nine”, and confusing “oh” and “no”), as well as the residual error.

# Auditory Scene Analysis

Al Bregman

A field like computer vision, but for audio.

Like vision, we start with raw data, and build models of the world.



Resynthesis is playing back the model.



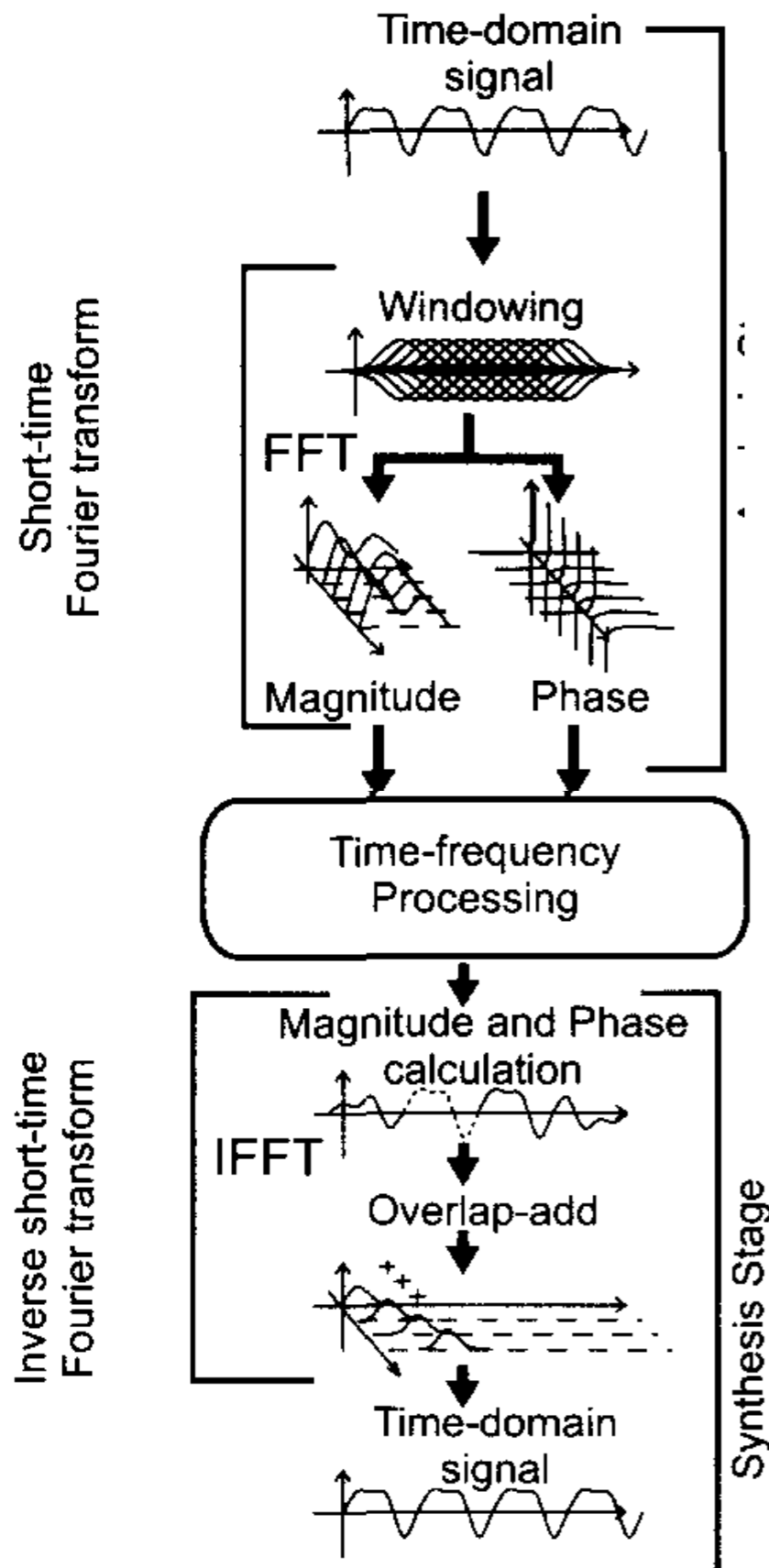
# Contrast with ...

---



# Audio Image Processing

“Audio Photoshop”  
Convert audio  
into a “spectral  
photo”.



Transform “photo” to be “better”  
(time-shift, pitch-shift, etc).

To recreate audio,  
“invert” photo-making  
process.

# Example: Phase Vocoder

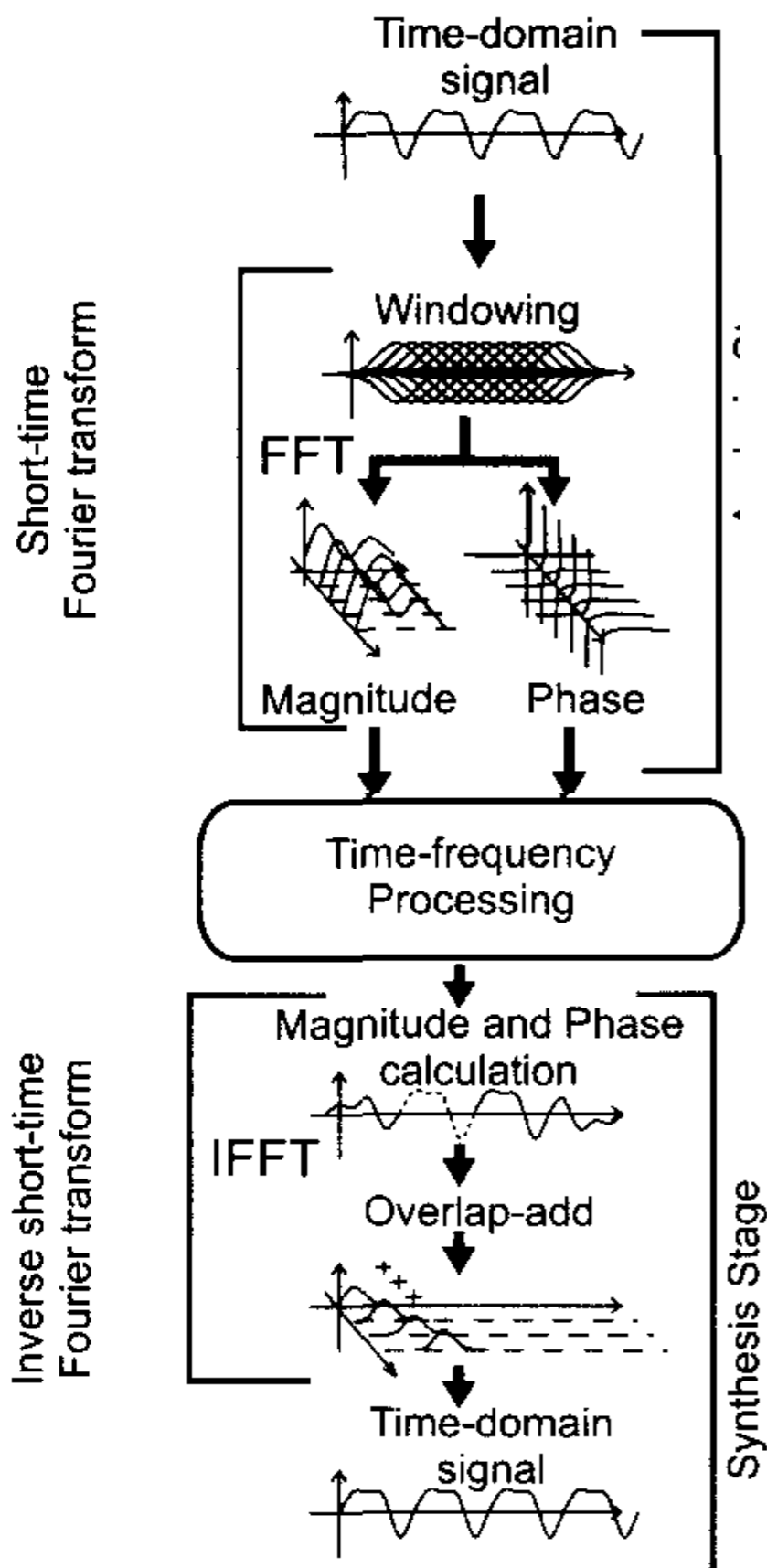
Original  
2X slower

## Advantages

No model to build:  
easier to design,  
faster to run.

## Disadvantages

Without a model,  
harder to maintain  
"common fate" and  
avoid artifacts.



# Admin: Progress Report Presentations

---

Progress Report Presentation	March 23 in class	A 10-15 minute presentation to the class, describing the current status of the project. Group projects should share presentation duties between all members. Audio demos of work in progress is encouraged. Primary purpose of presentation is to solicit feedback from the audience.	15 percent
------------------------------	-------------------	---	------------

**Cynthia**

**Psyche  
(Apr 6)**

**Jeremy**

**Bradley**

**Carlos**

**Eric**

**If you are enrolled (or are auditing and doing a project) and not on the list, let us know!**