EECS 225D
Audio Signal Processing in Humans and Machines

Lecture 15 – Pitch

2012-3-12

Professor Nelson Morgan

today's lecture by John Lazzaro

www.icsi.berkeley.edu/eecs225d/

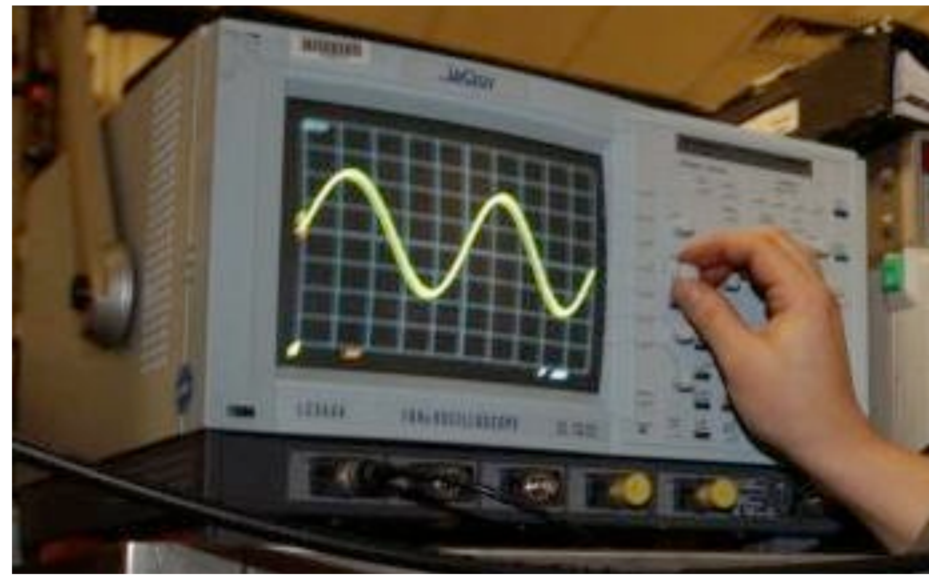
A_3 = 220 Hz A_4 = 440 Hz Play.

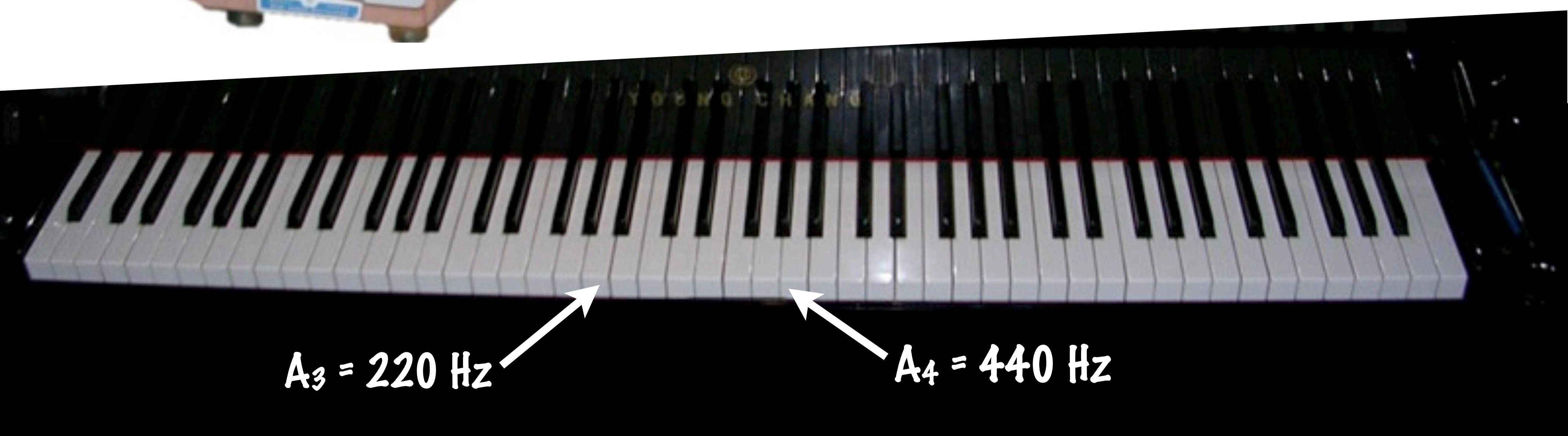
Today's lecture: Pitch

- Basic concepts in pitch detection
- Practical issues in pitch tracking
- The quiz ...

Musical pitch: an experimental definition

Pitch (unit: Hz). **The frequency of a sine wave whose pitch is heard to be the same as a played note.**





Timbre and Pitch

Why are the timbres different?

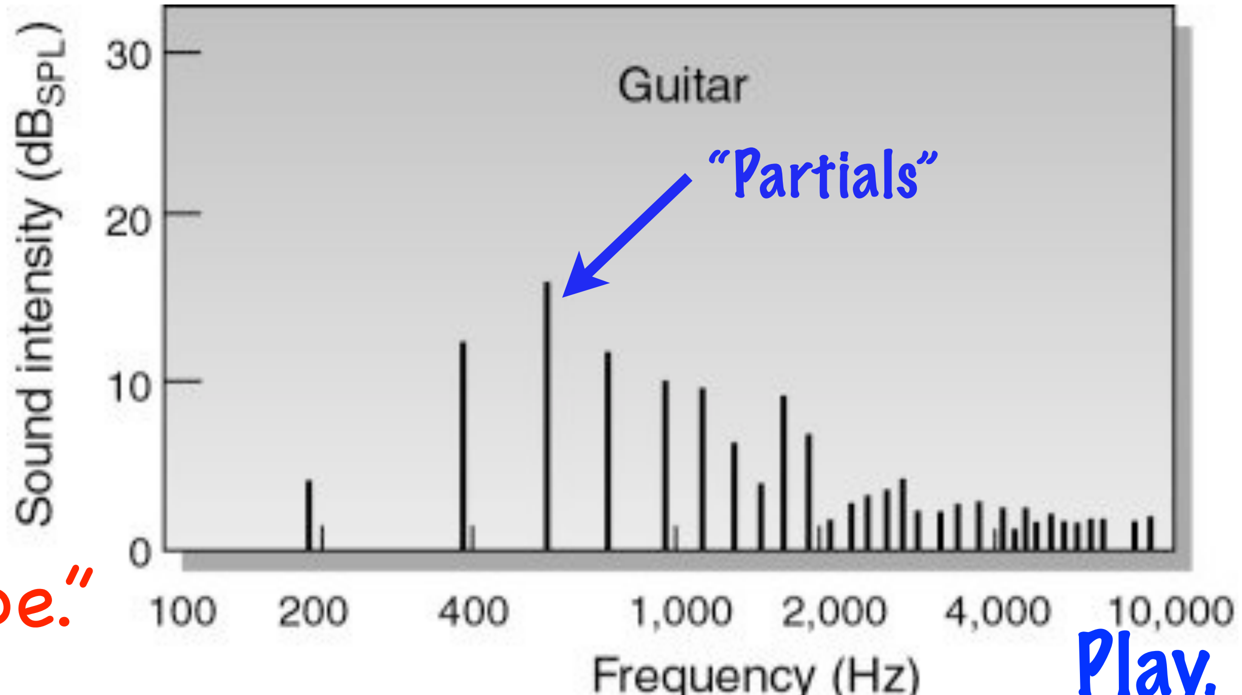
Contributing factor: Partial heights differ, and evolve differently over time. "Spectral Shape."

Why are both sounds pitched?

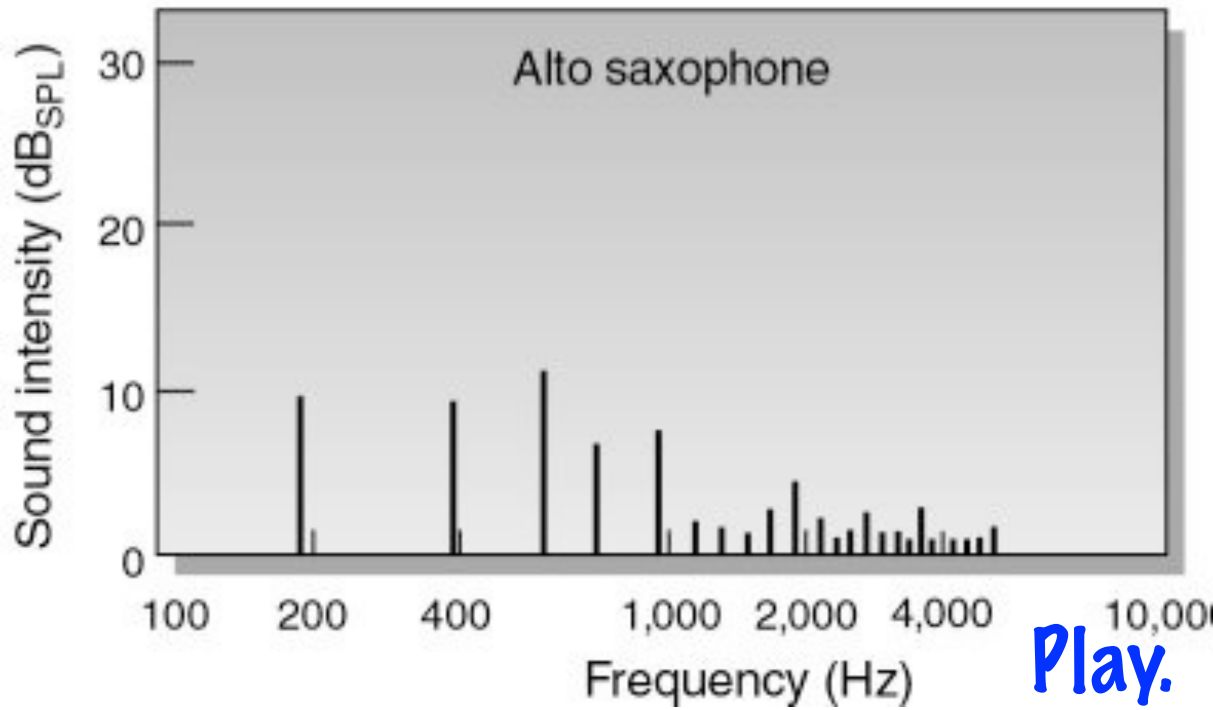
Why is the pitch the same?

Frequency placement of partials share a common structure.

Same pitch, different timbre

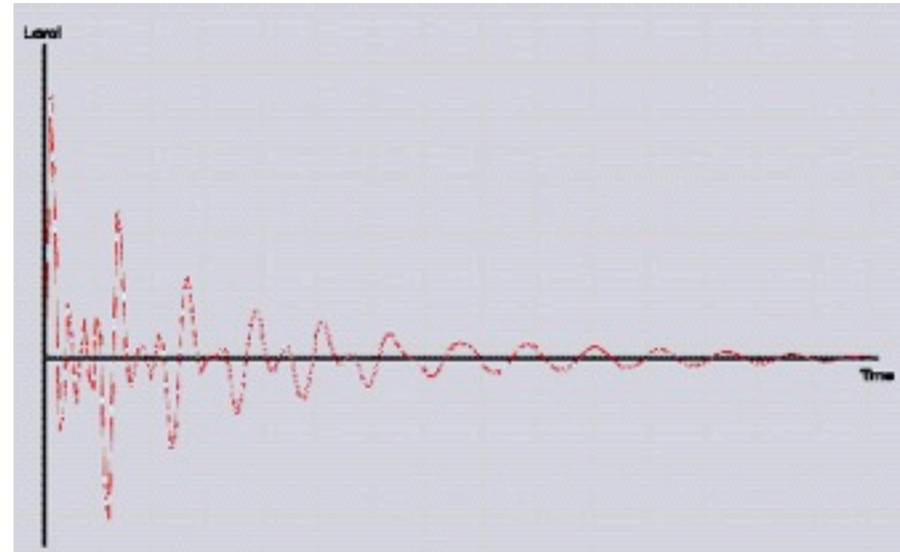


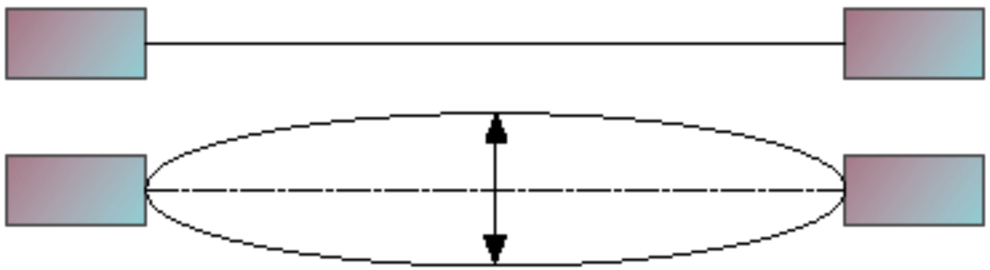
Play.

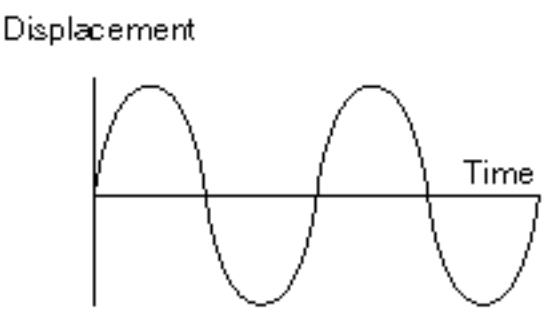


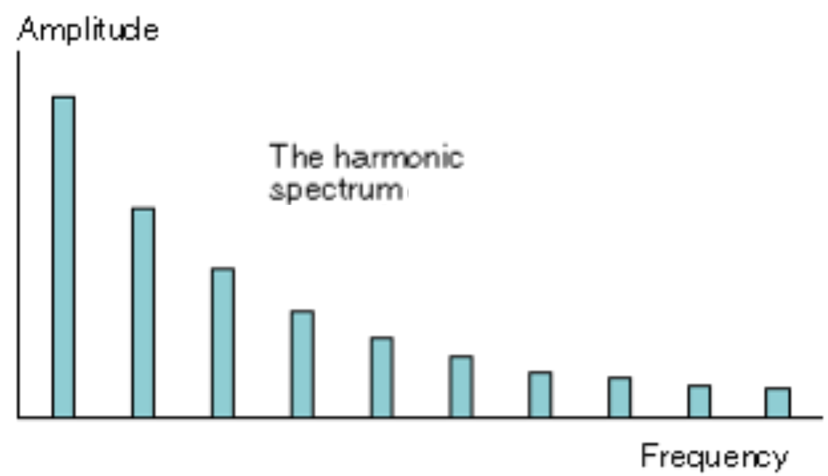
Play.

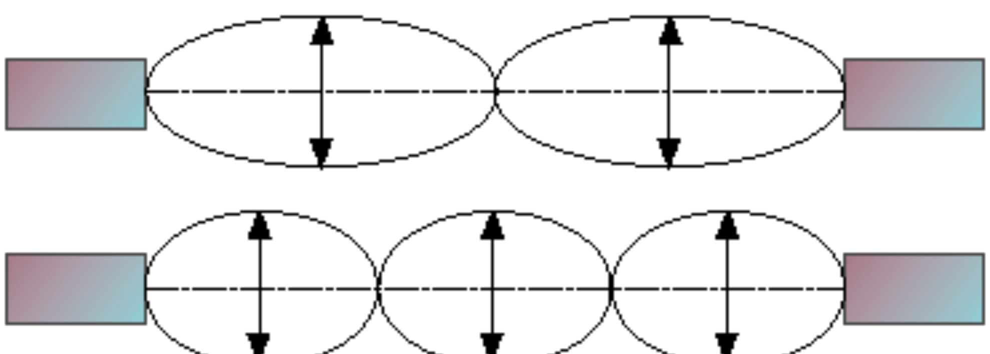
Example by Kenneth Steele, Appalachian State

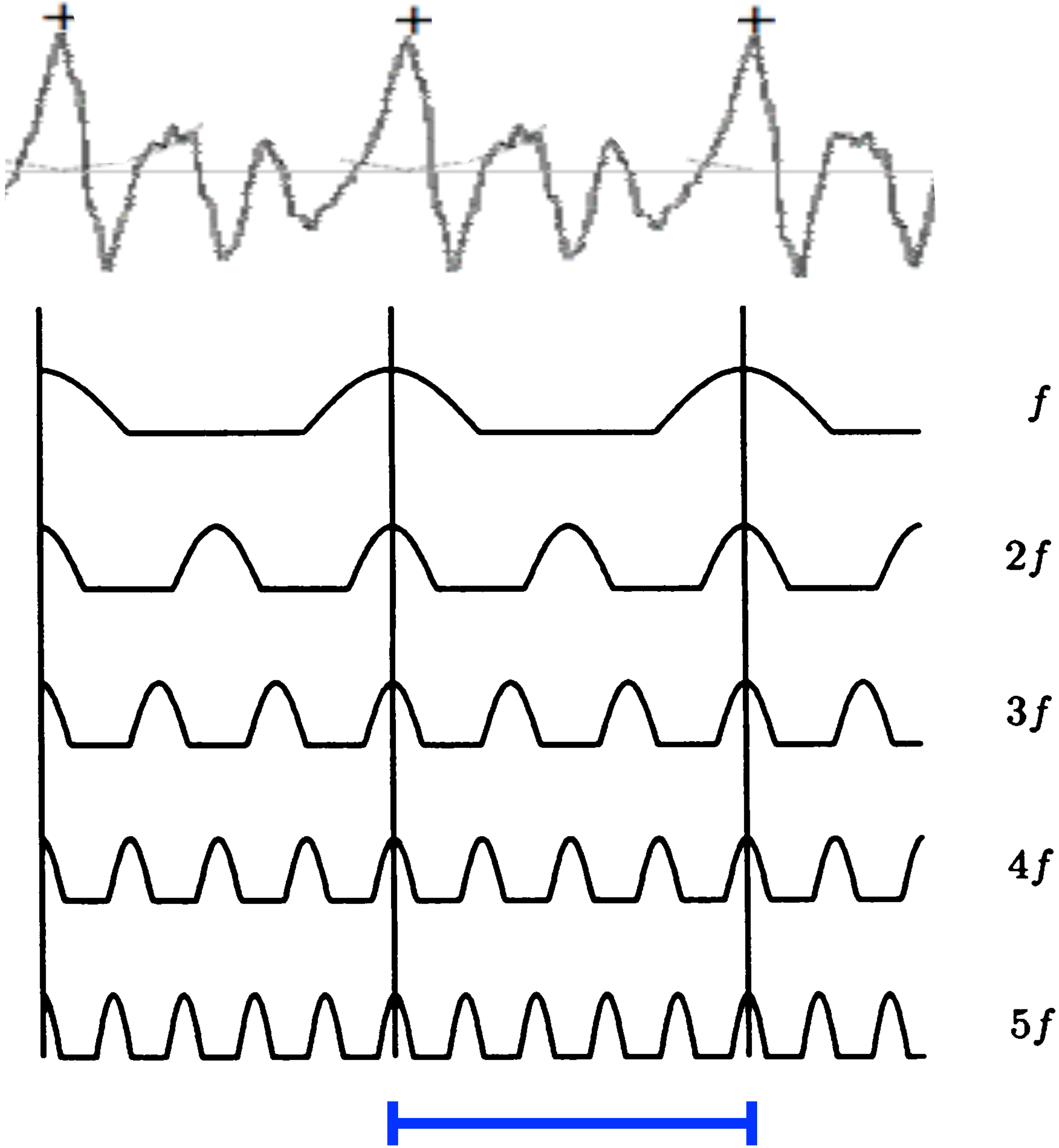





Displacement
Time


Amplitude
The harmonic spectrum
Frequency



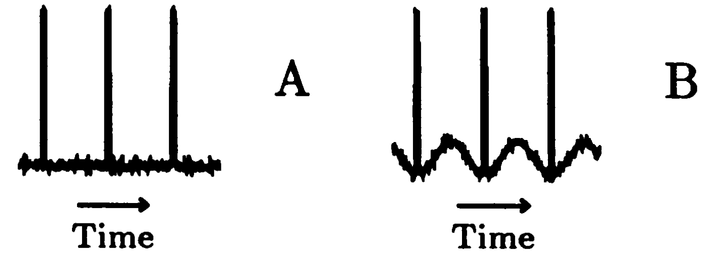


Summed waveform repeats at pitch frequency.

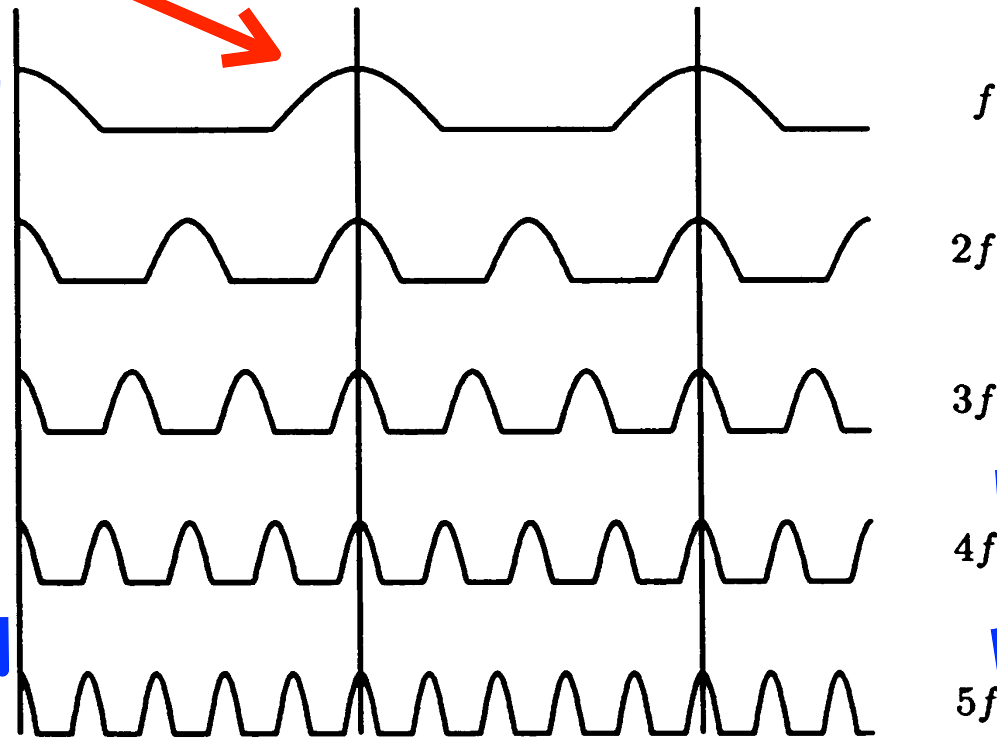
Frequencies of partials are integer multiples of an underlying fundamental.

Pitch Period = 1 / (Pitch Frequency)

Caveats: First partial not necessary to detect pitch - A and B → are heard with same pitch.

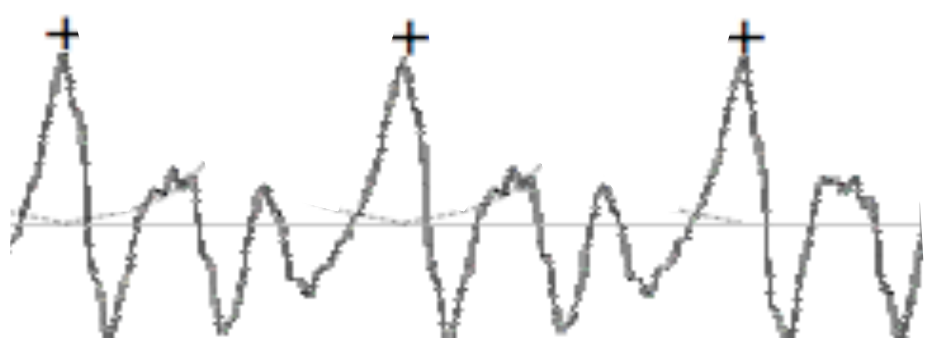


Relative phases of partials need not be aligned - any phase relation yields a strong pitch.



Sounds whose partials are not quite integer - related still yield a sense of pitch -

Thus ... repeating shape may be subtle to detect directly.

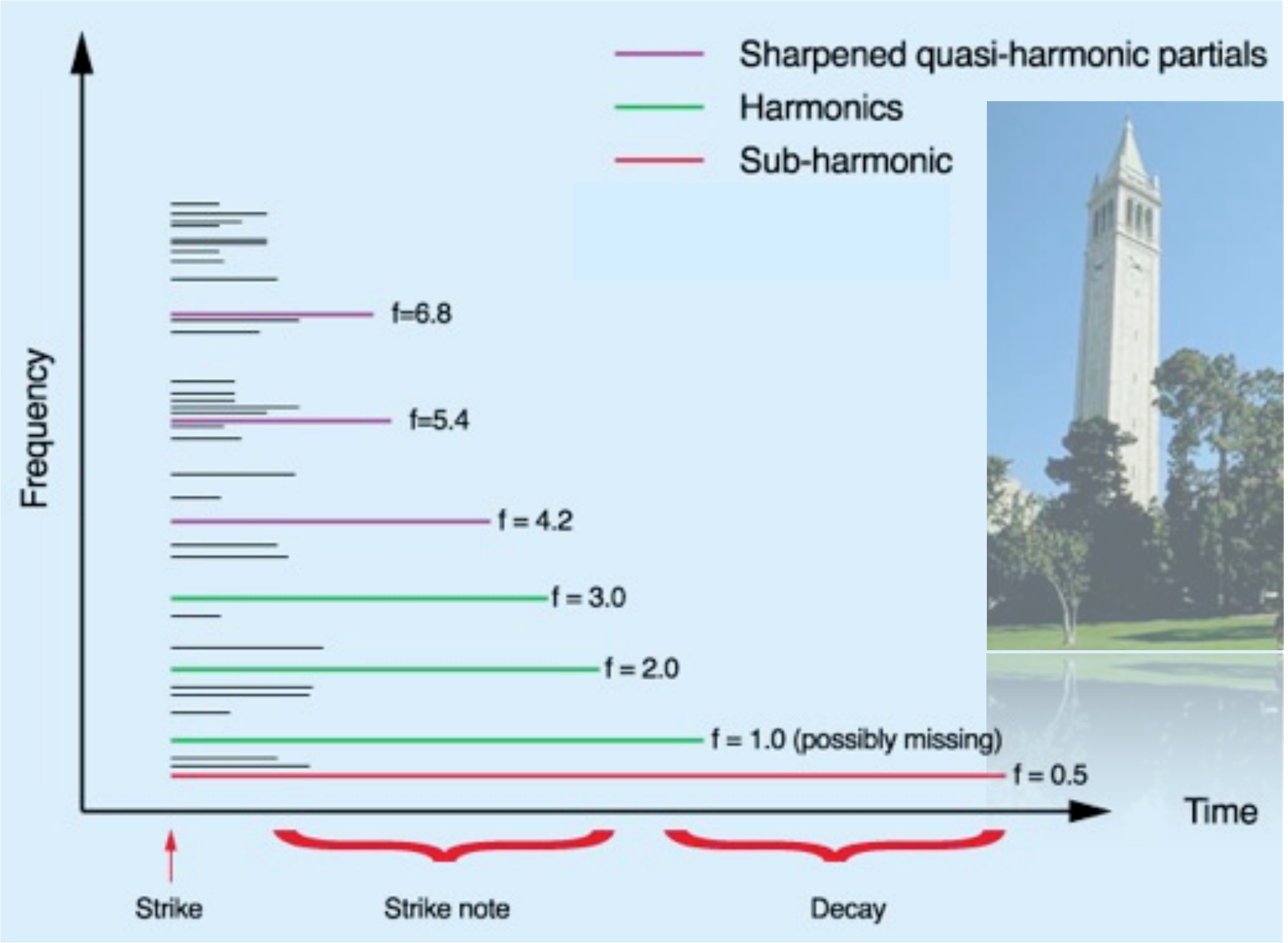


Bells

Lowest partials are exact integers, but higher partials are quasi-harmonic (4.2, 5.4, 6.8).

We still hear the bells as having a definite pitch.

Play.

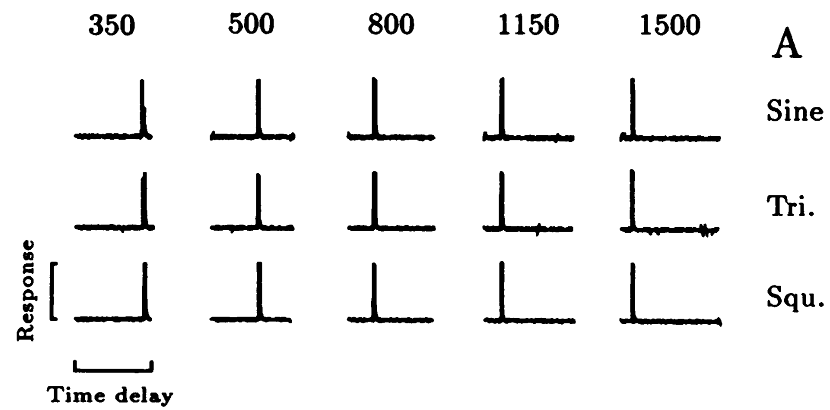


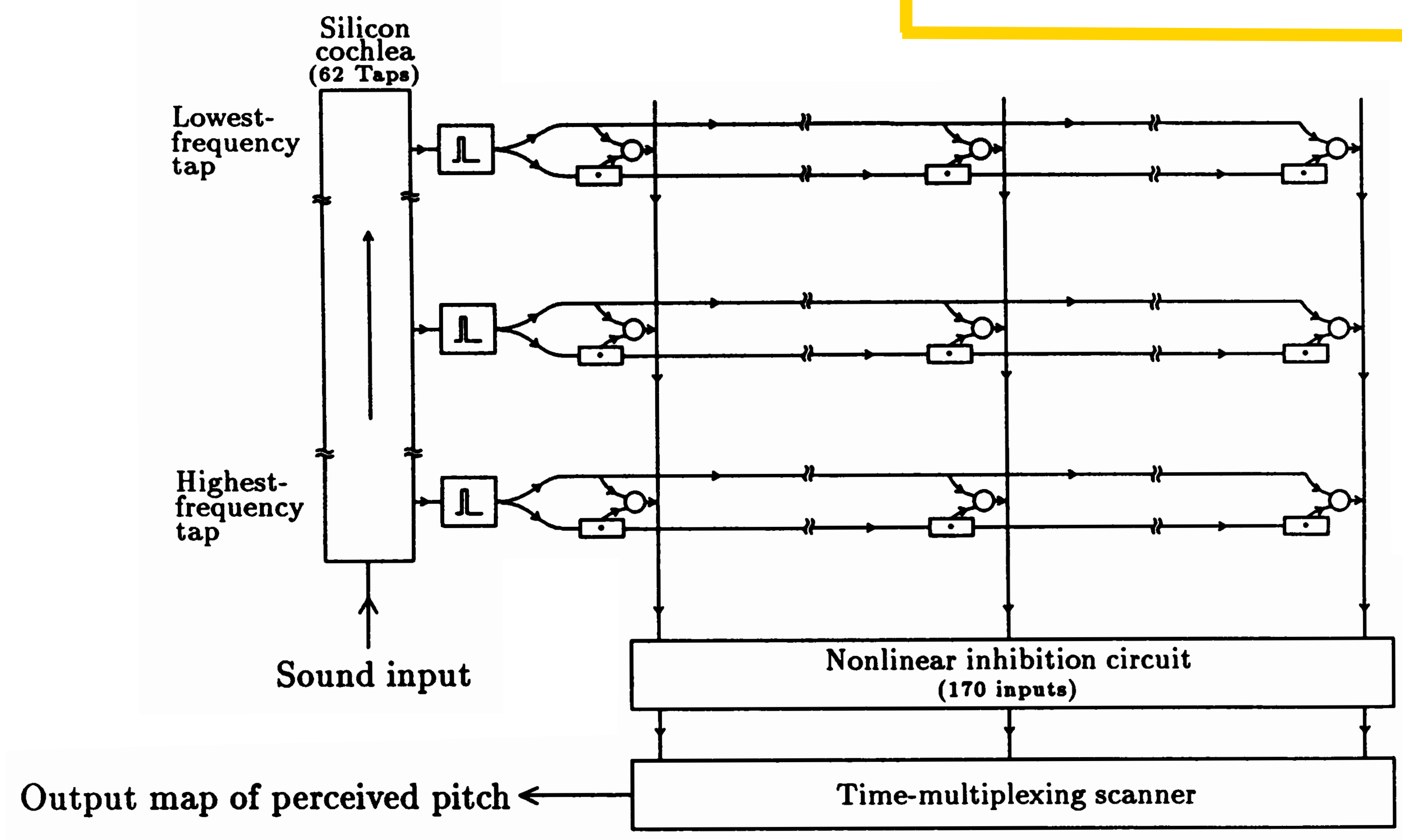
But bell "chords" often sound atonal.

FROM SOUND ON SOUND, SYNTH SECRETS, AUG 02

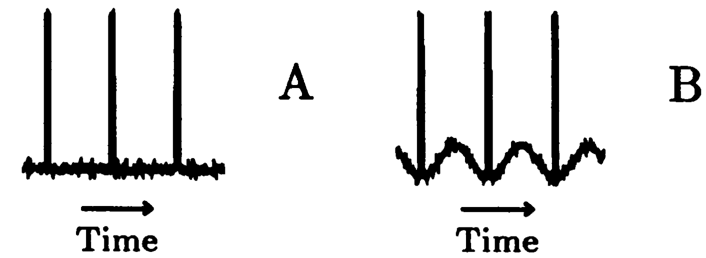
Licklider Pitch Model

Autocorrelate filtered versions of the audio waveform.

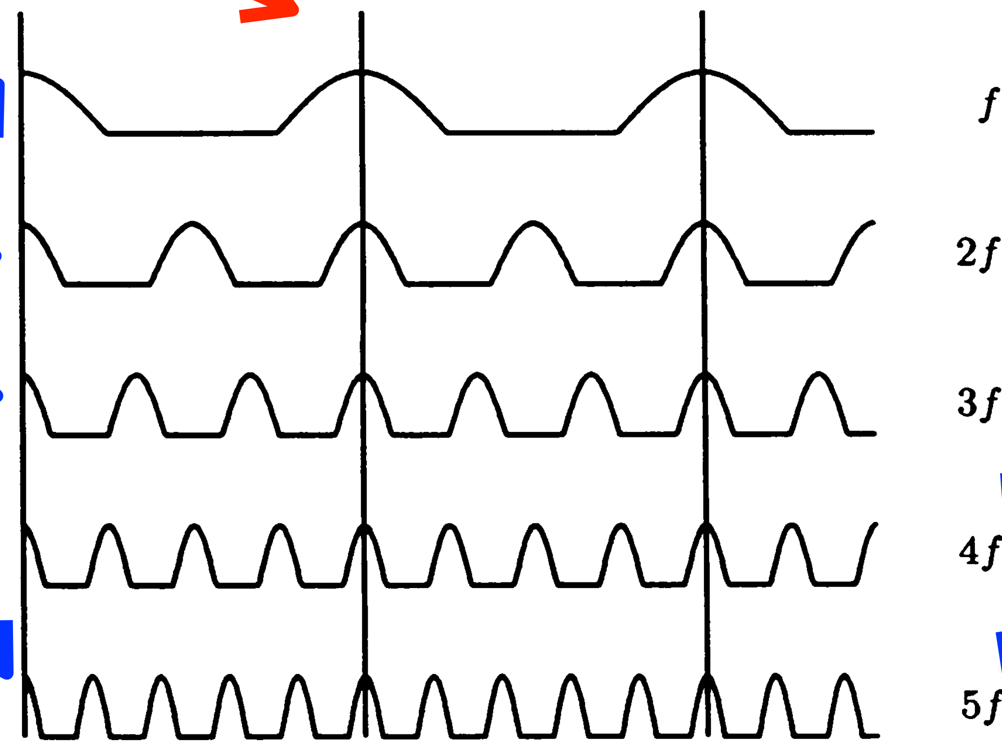




Recall: First partial not necessary to detect pitch - A and B are heard with same pitch.



Relative phases of partials need not be aligned - any phase relation yields a strong pitch.



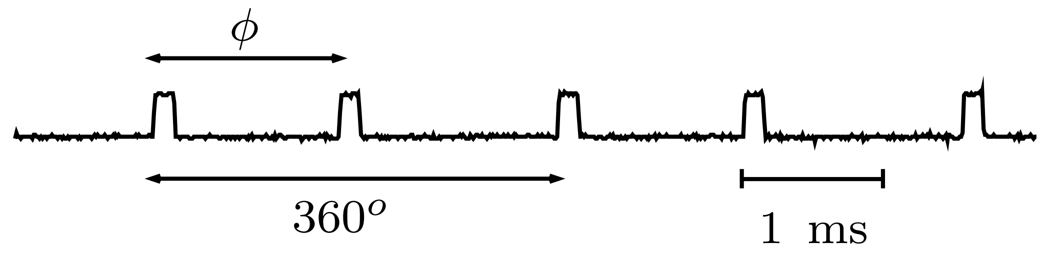
Sounds whose partials are not quite integer - related still yield a sense of pitch -

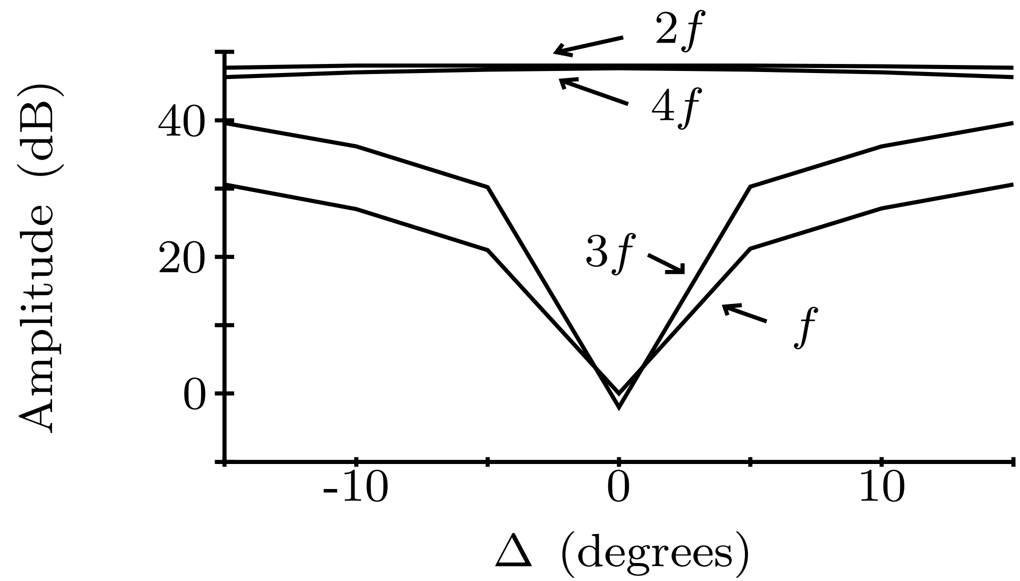
Autocorrelation model addresses all of these issues.

Seebeck's Siren

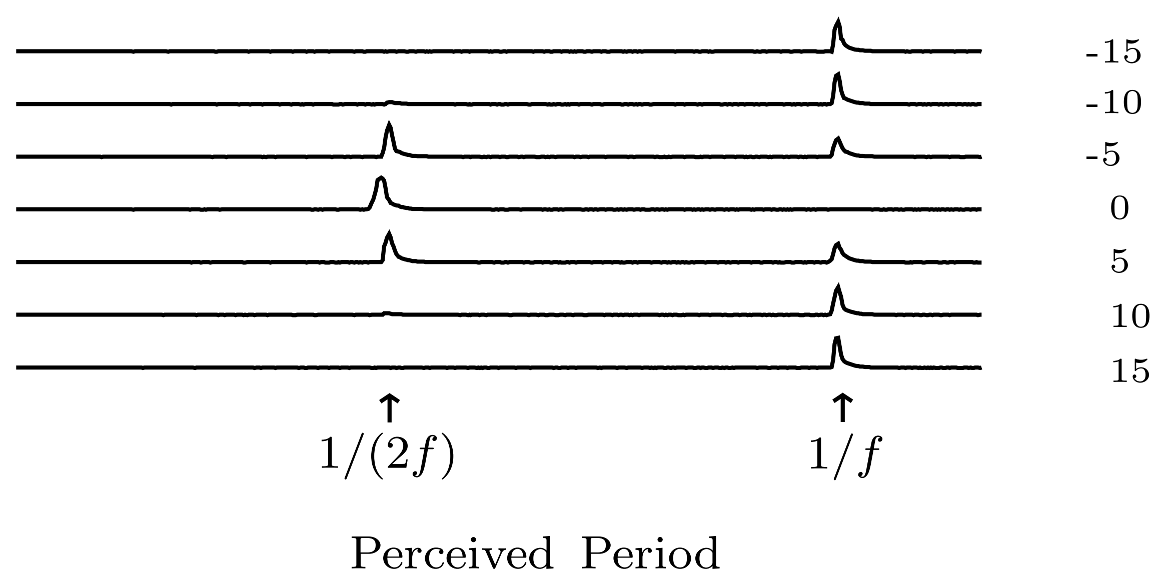
When ϕ is offset slightly from 180°, the pitch **drops** one octave. Why? The repetition period of the waveform **doubles**.

Autocorrelation "hears" this signal the **same way humans do**.

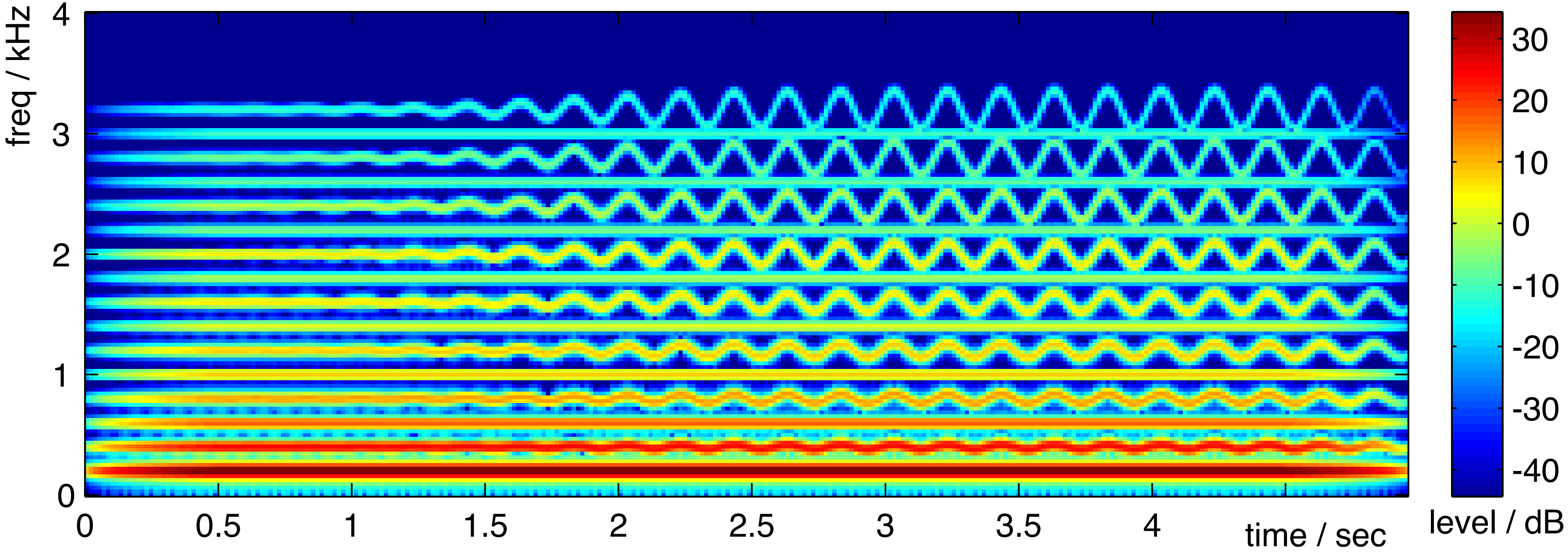




Play.



Related: Reynolds/McAdams Oboe



Adding vibrato to even partials "creates" a second sound whose pitch is raised by one octave (spectrum by Dan Ellis).

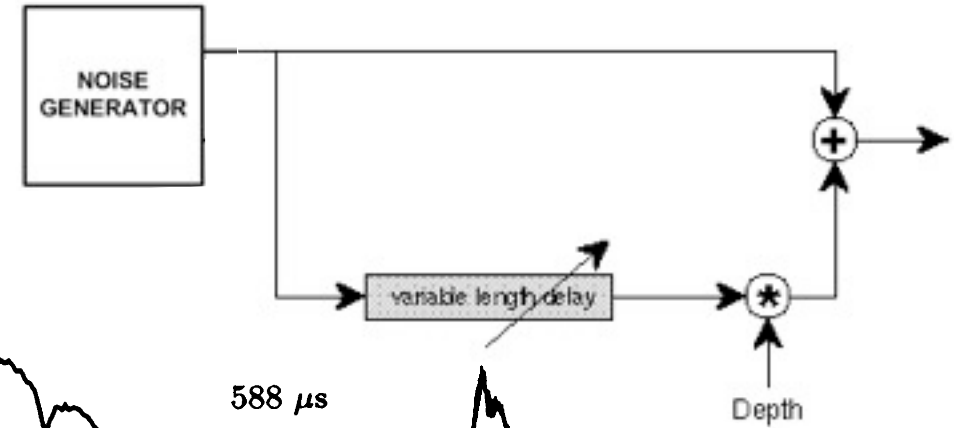
Play

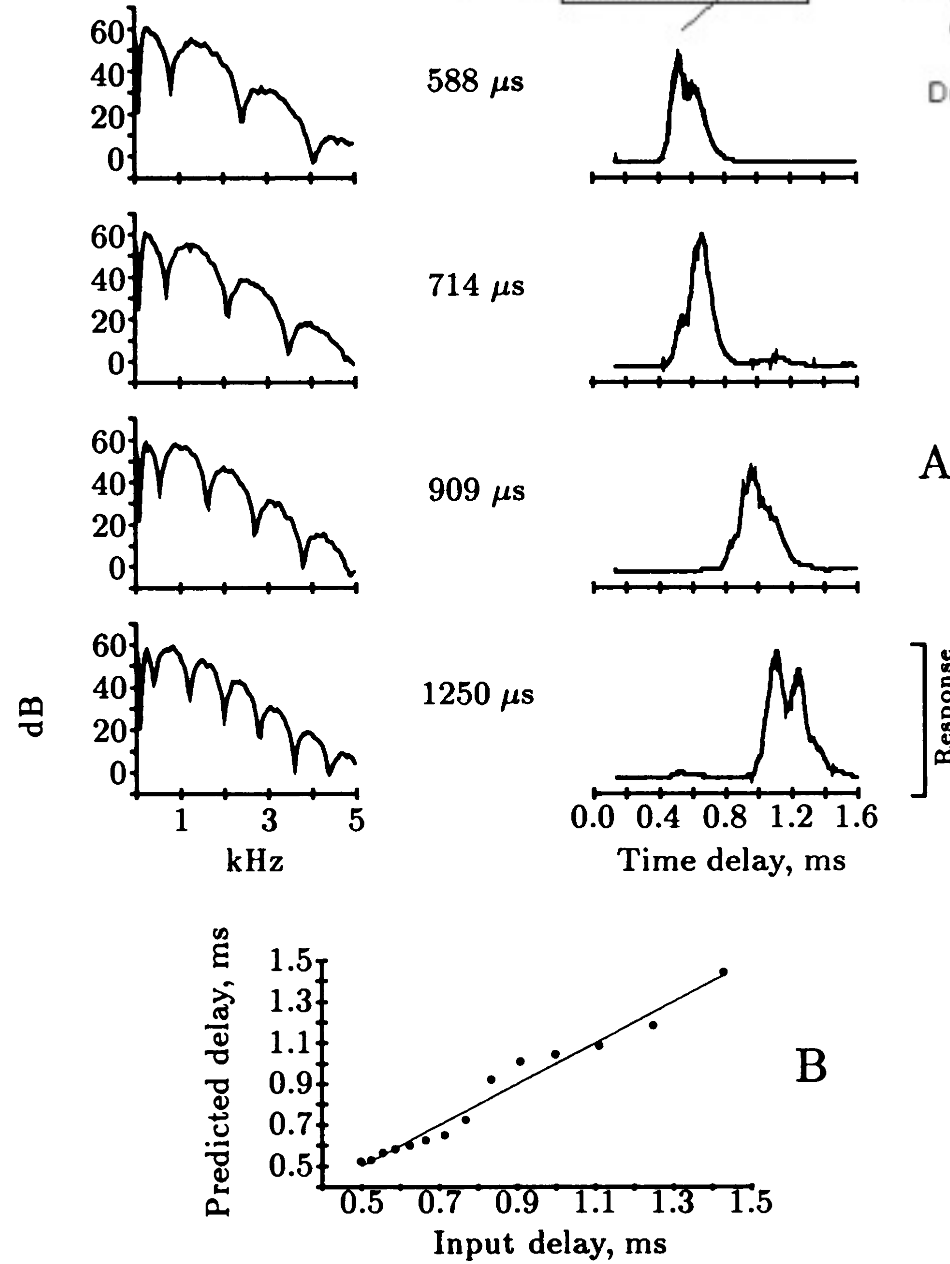
Comb-Filtered Noise

We hear a weak pitch in comb-filtered noise, corresponding to the period of the comb delay.

This is unusual because the waveform of each period is unique.

Autocorrelation "hears" the same pitch humans perceive.





Play.

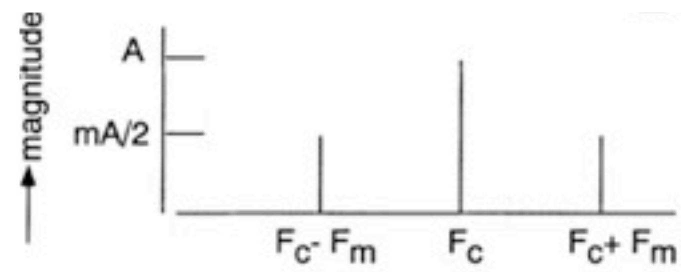
Amplitude Modulation

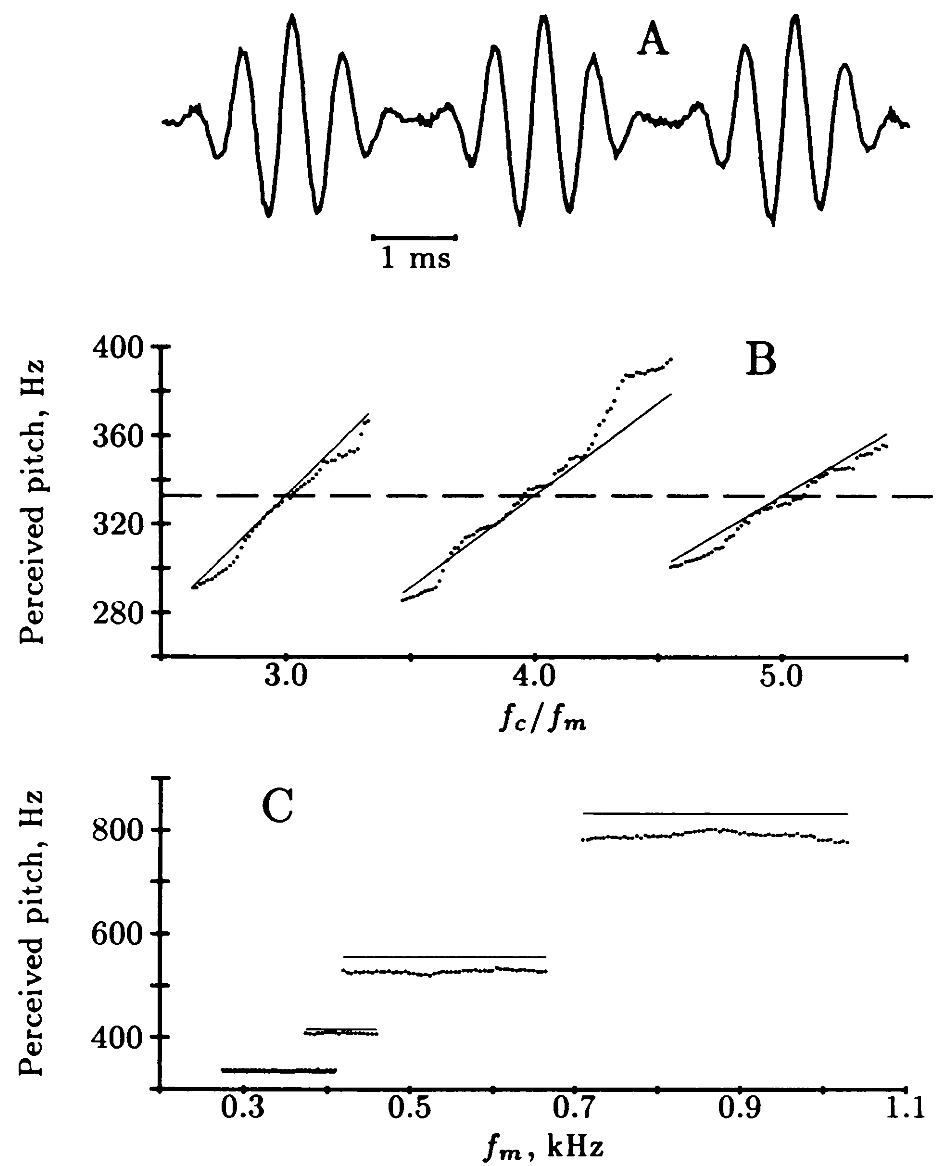
This sound has components at $F_c - F_m$, F_c, and $F_c + F_m$.

To a first order, the pitch we hear is F_m (the repetition frequency).

Autocorrelation matches this first-order result.

But fails on second-order phenomena.
(slope of pitch for small changes in F_c , F_m).

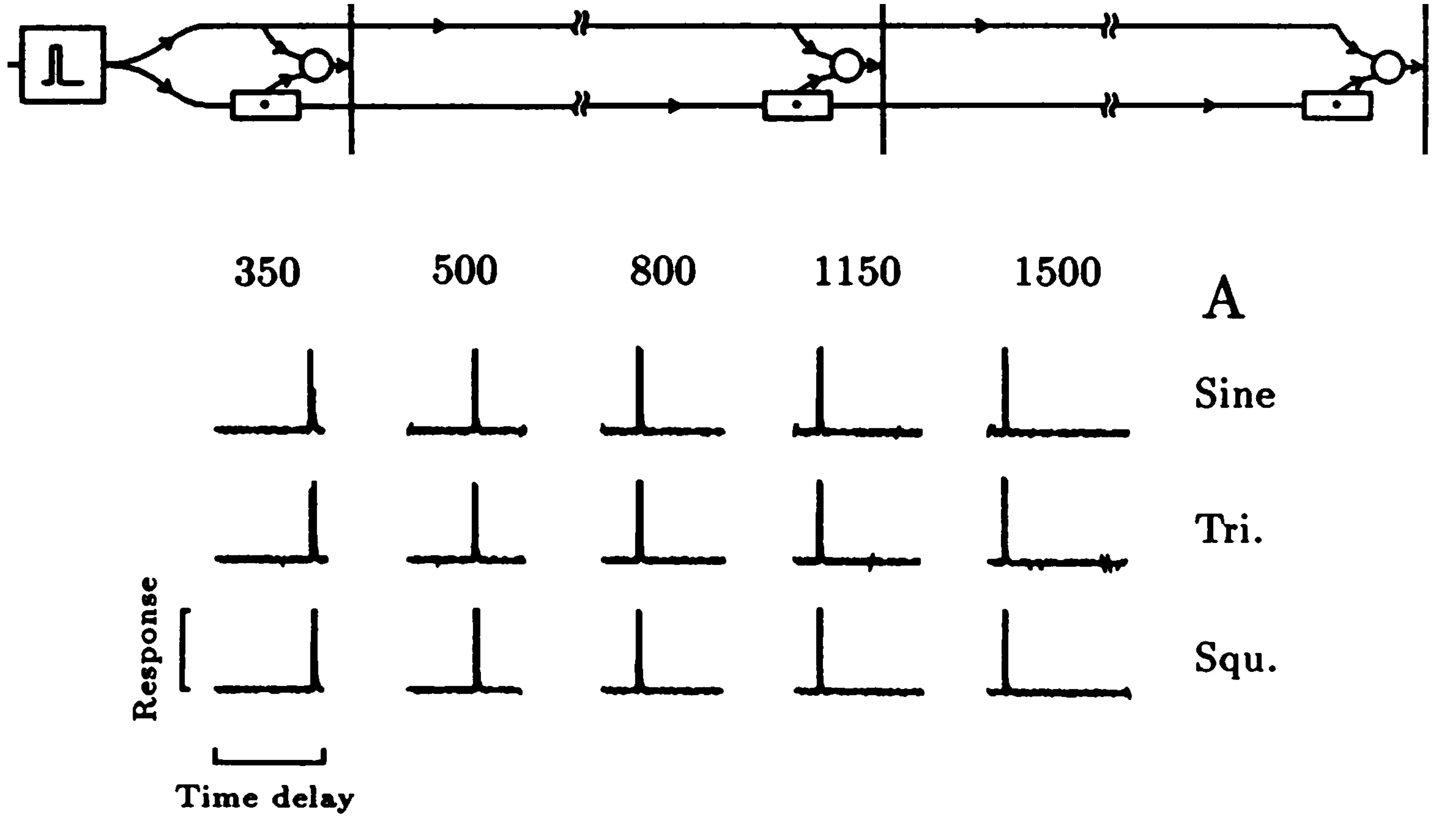




Today's lecture: Pitch

- Basic concepts in pitch detection
- **Practical issues in pitch tracking**
- The quiz ...

Modern engineering pitch trackers are based on multi-tap autocorrelation (or similar operators).



Algorithm enhancements mostly take the form of pre-processing the audio input or post-processing the auto-correlation tap outputs, to better handle the "difficult" pitch signals we see in real-world engineering applications.

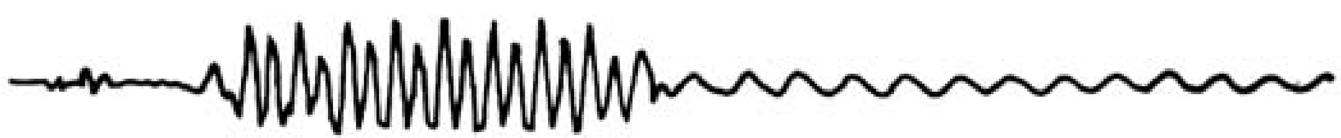
Real-World Issues

Dynamic range of pitch

(*a*)

Pitch variations in time

(*b*)

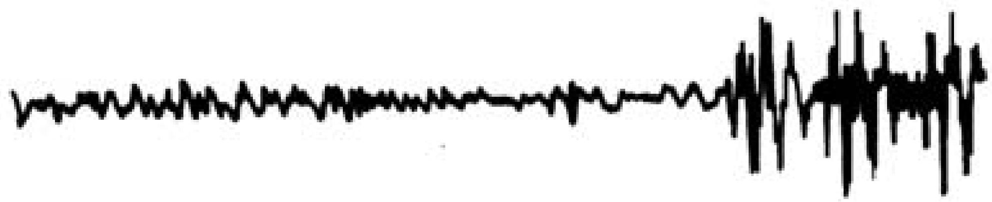


Vocal tract variations in time

(*c*)

10 ms

Real-World Issues

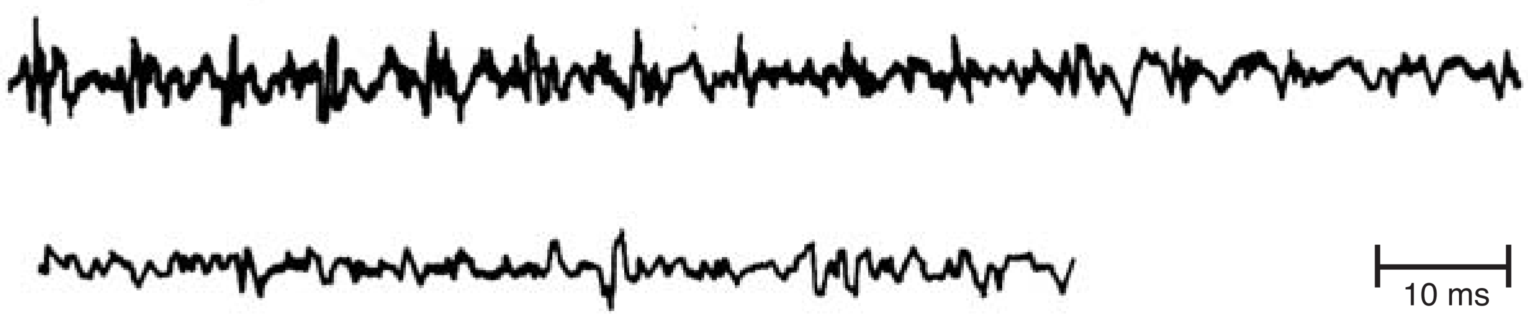


Voiced-unvoiced transition

(*d*)

Telephone speech

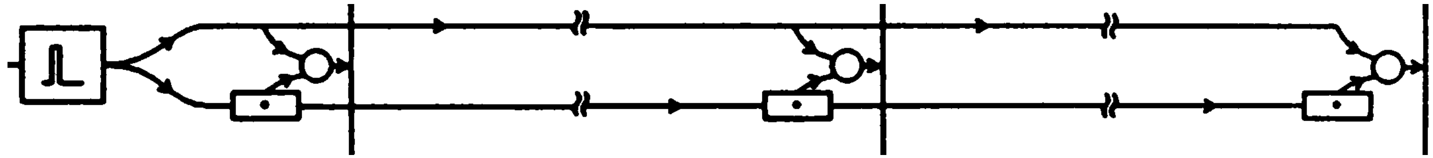
(*e*)

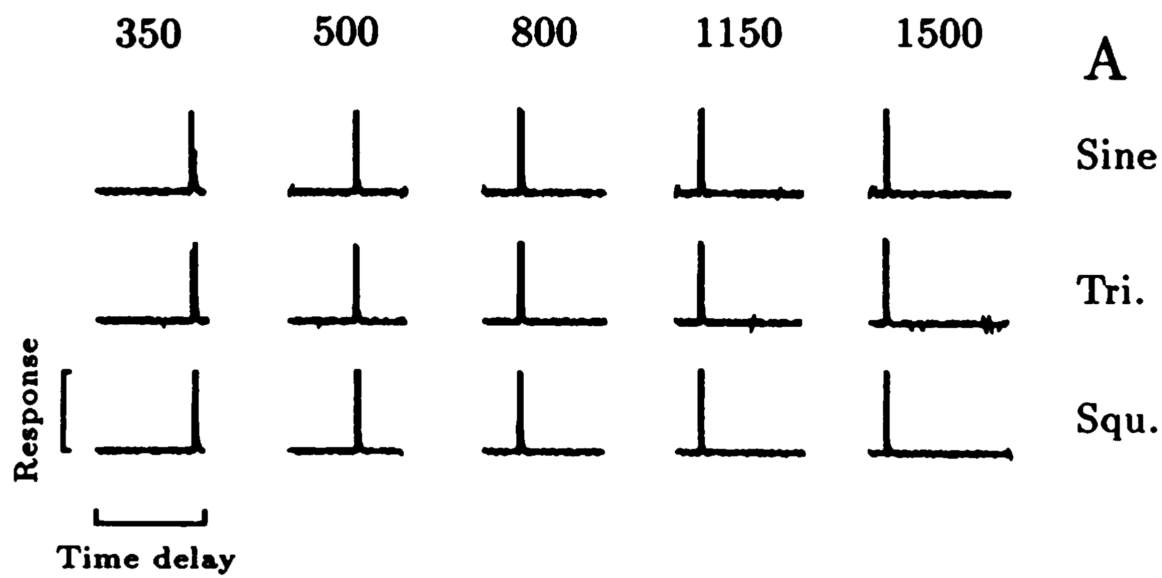


Acoustic noise background

(*f*)

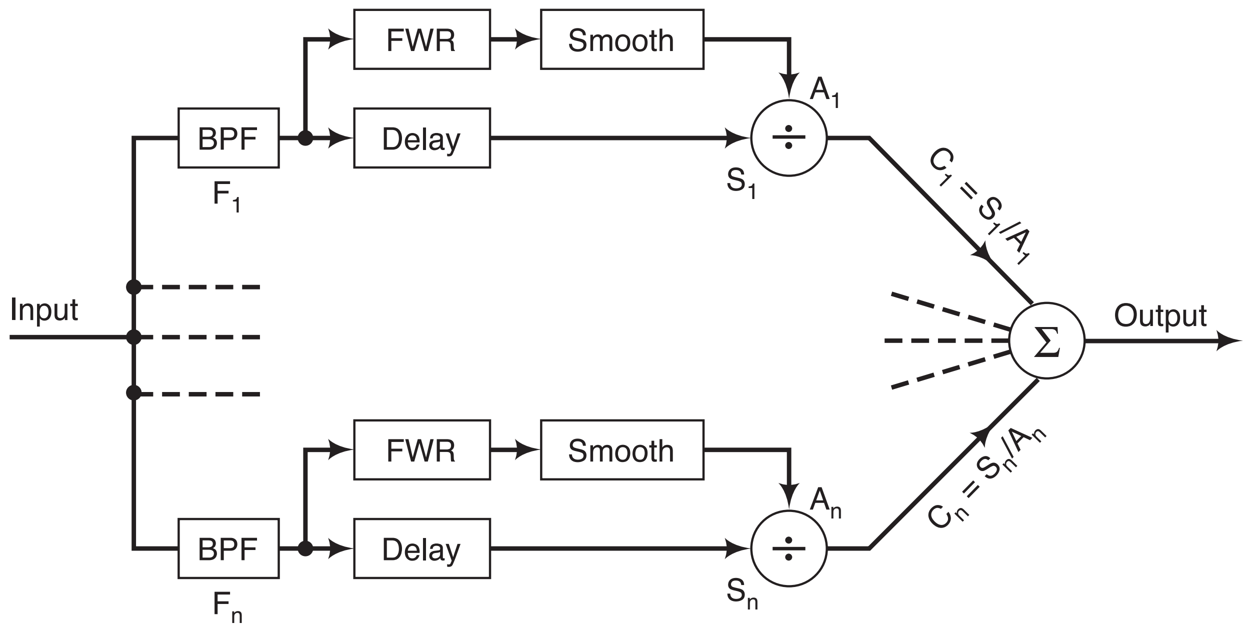
Modern engineering pitch trackers are based on multi-tap autocorrelation (or similar operators).

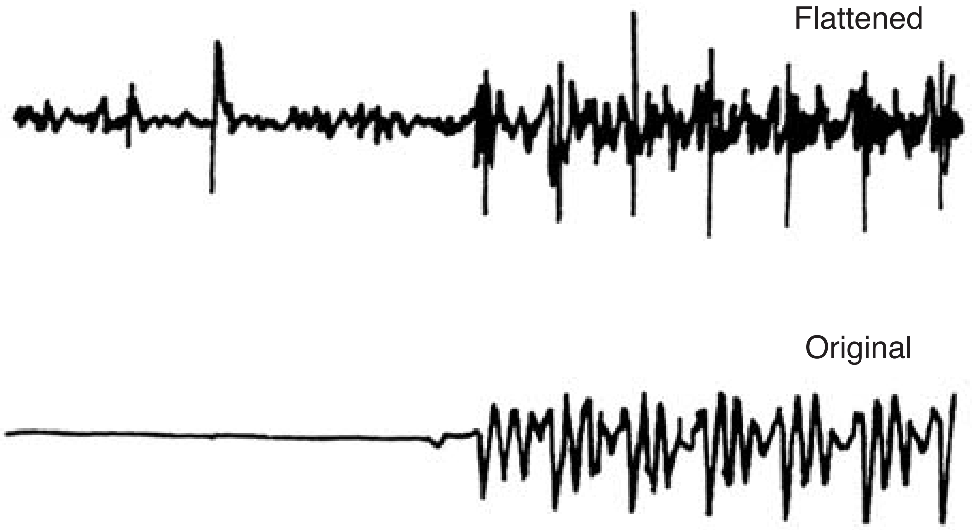




Algorithm enhancements mostly take the form of pre-processing the audio input or post-processing the auto-correlation tap outputs, to better handle the "difficult" pitch signals we see in real-world engineering applications.

Pre-Processing : Spectral Flattening





Pre-Processing : Low-pass Filtering

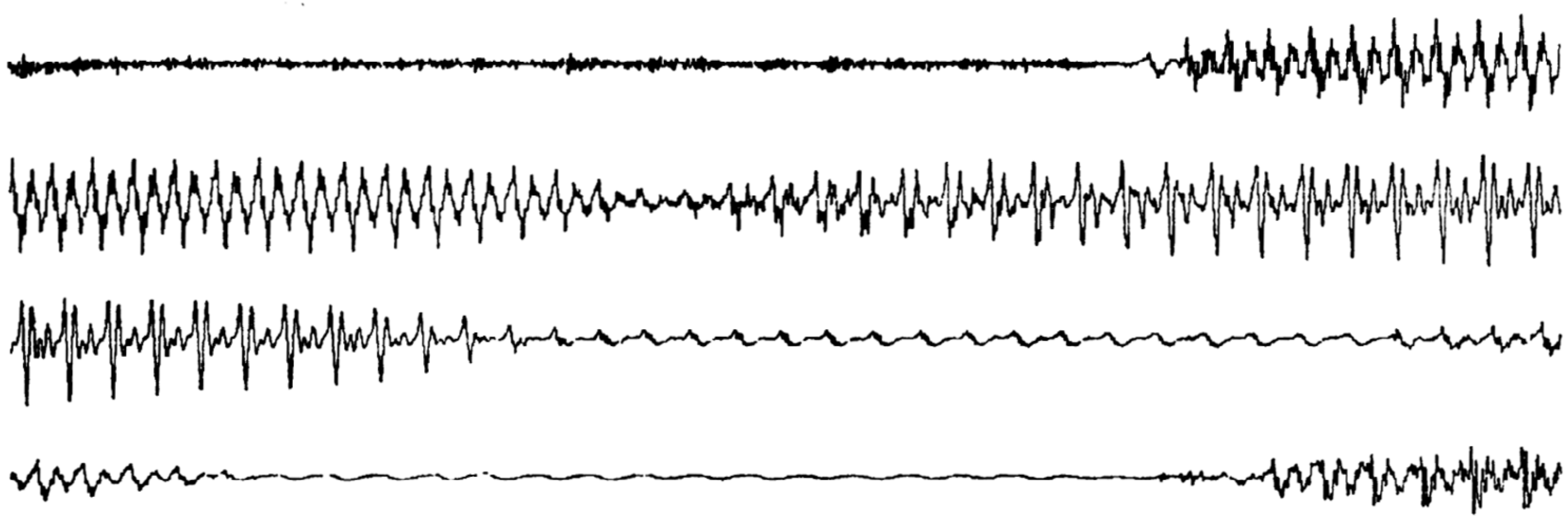


FIGURE 31.5 Full-band speech signal.

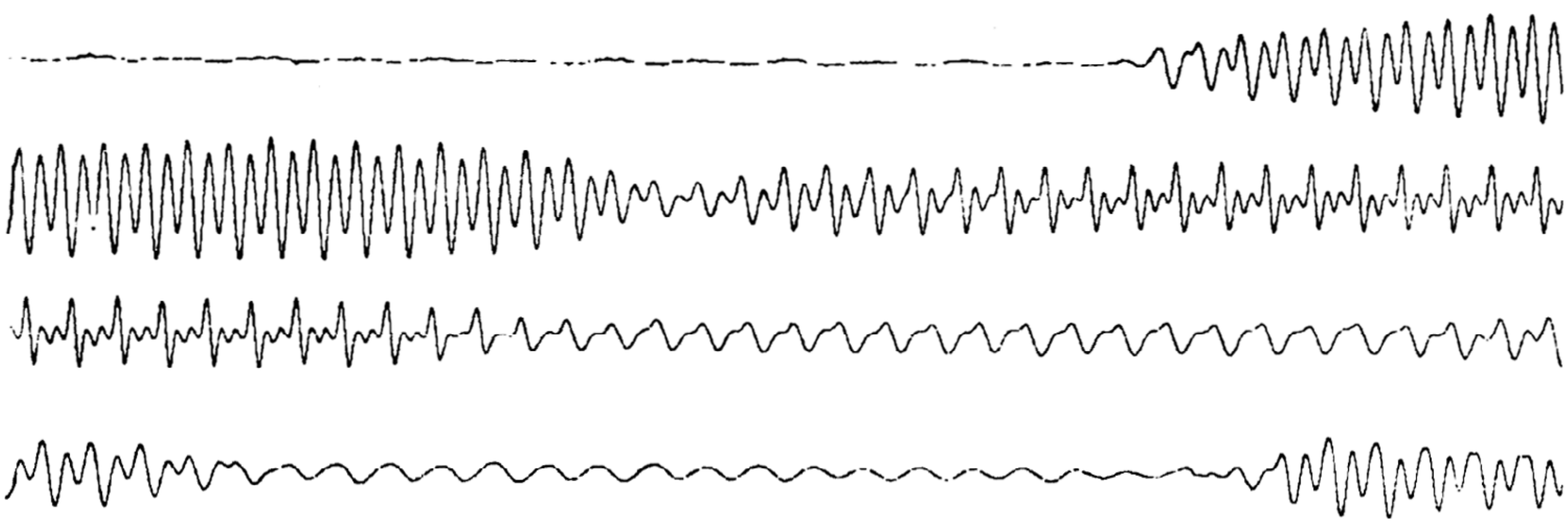
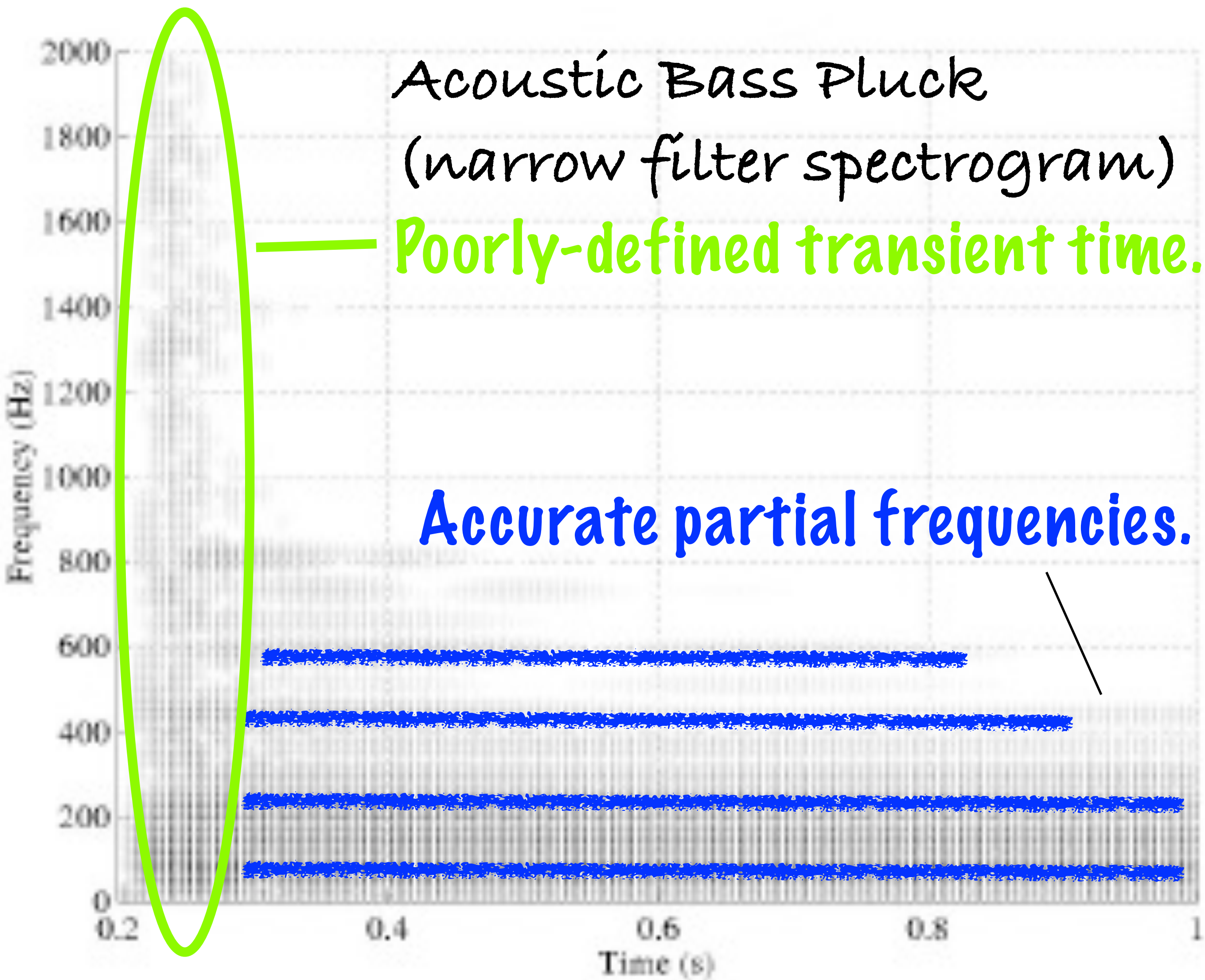


FIGURE 31.6 Low-pass filtered speech signal.

Time/Frequency Tradeoffs

Narrow linear filters are **slow**

Fast linear filters are **wide**

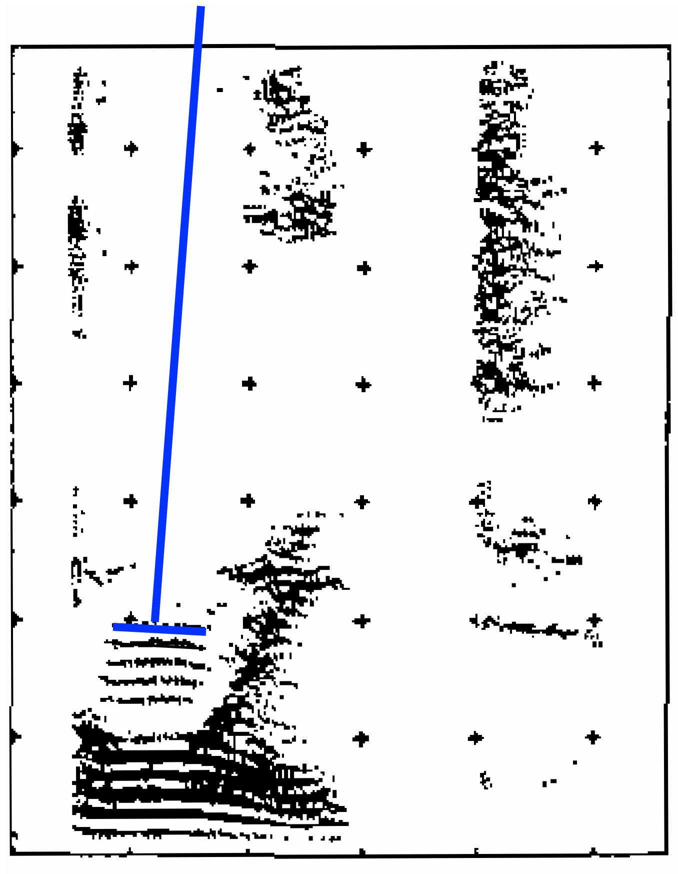


Filters: Wide vs Narrow

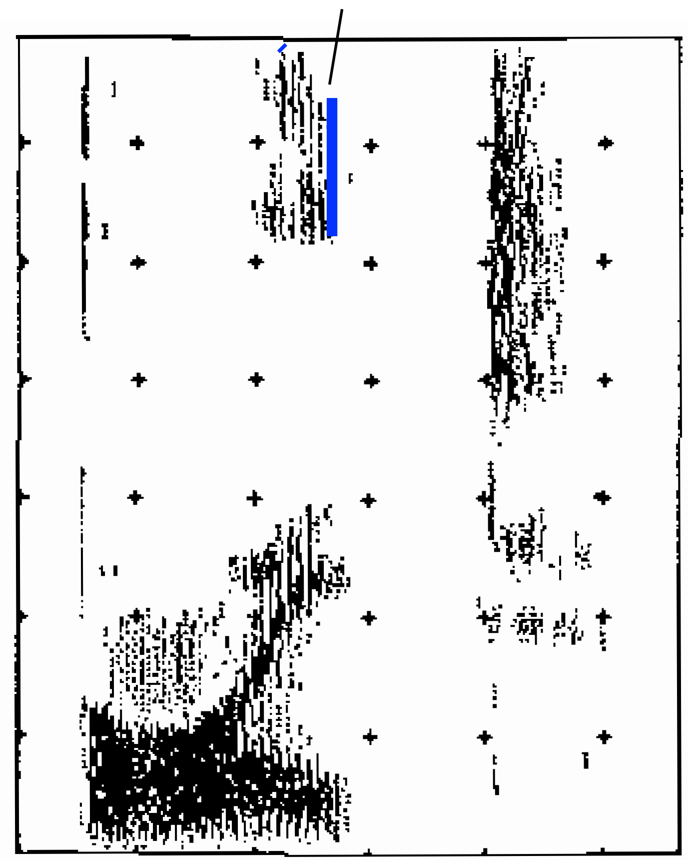
Spoken word "boyt" processed by two filterbanks.

Accurate pitch harmonics.

Good glottal pulse timing.



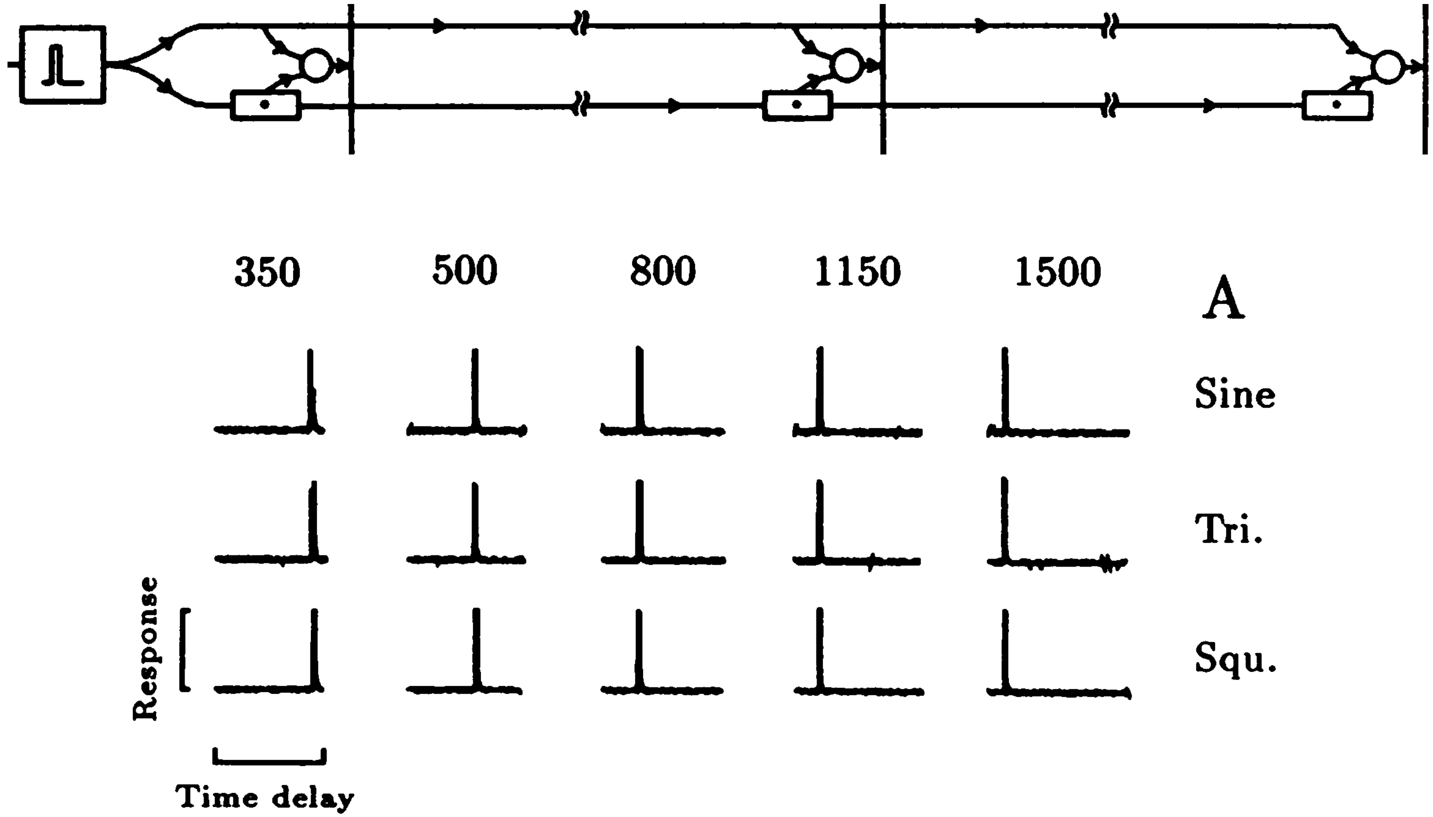
Narrow filters



Fast filters

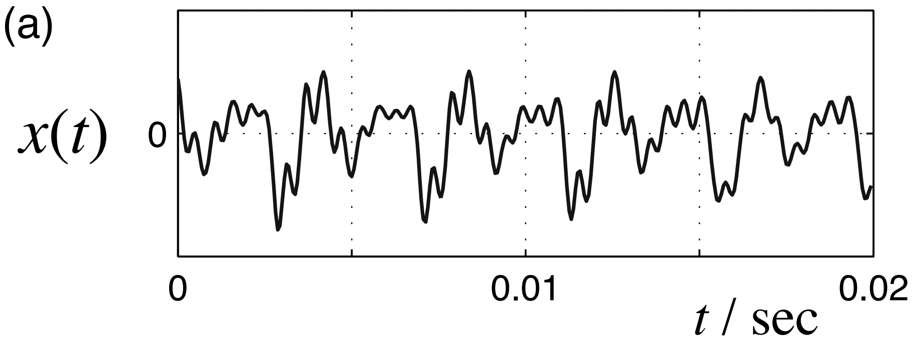
Data from Hong Leung and Victor Zue

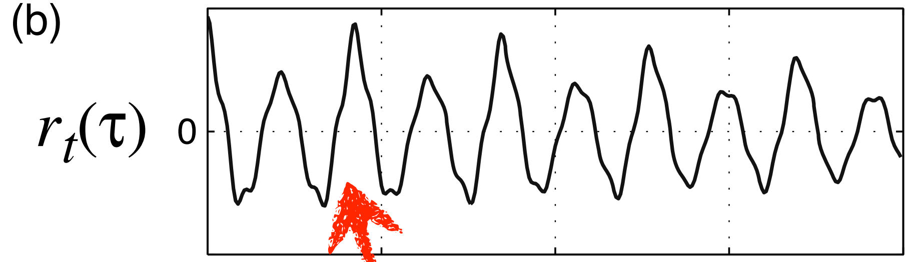
Modern engineering pitch trackers are based on multi-tap autocorrelation (or similar operators).



Algorithm enhancements mostly take the form of pre-processing the audio input or post-processing the auto-correlation tap outputs, to better handle the "difficult" pitch signals we see in real-world engineering applications.

Post-Processing : Yin Normalization

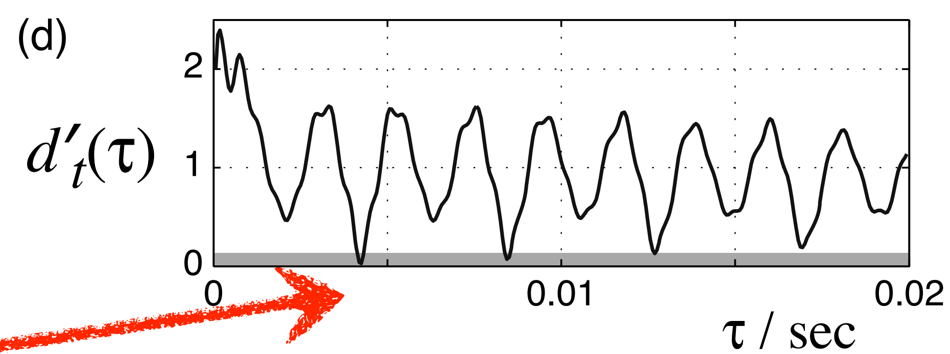




Female voiced speech, time domain waveform

Raw autocorrelation. Requires "peak" picking.

Yin normalization: autocorrelation peaks fall in a small threshold around "0"



Today's lecture: Pitch

- Basic concepts in pitch detection
- Practical issues in pitch tracking
- **The quiz ...**