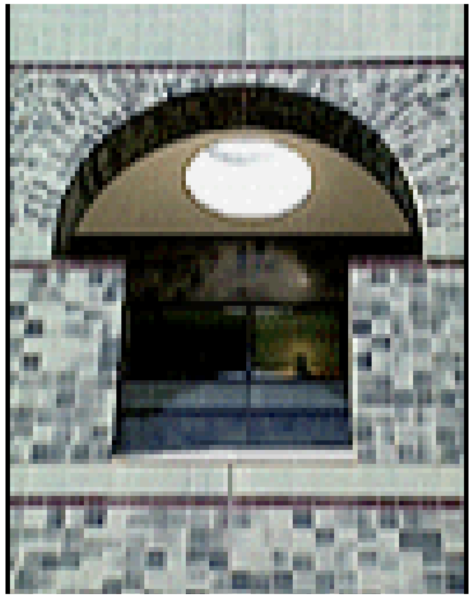
Music 209
Advanced Topics in Computer Music
Lecture 7 – Database Descriptors

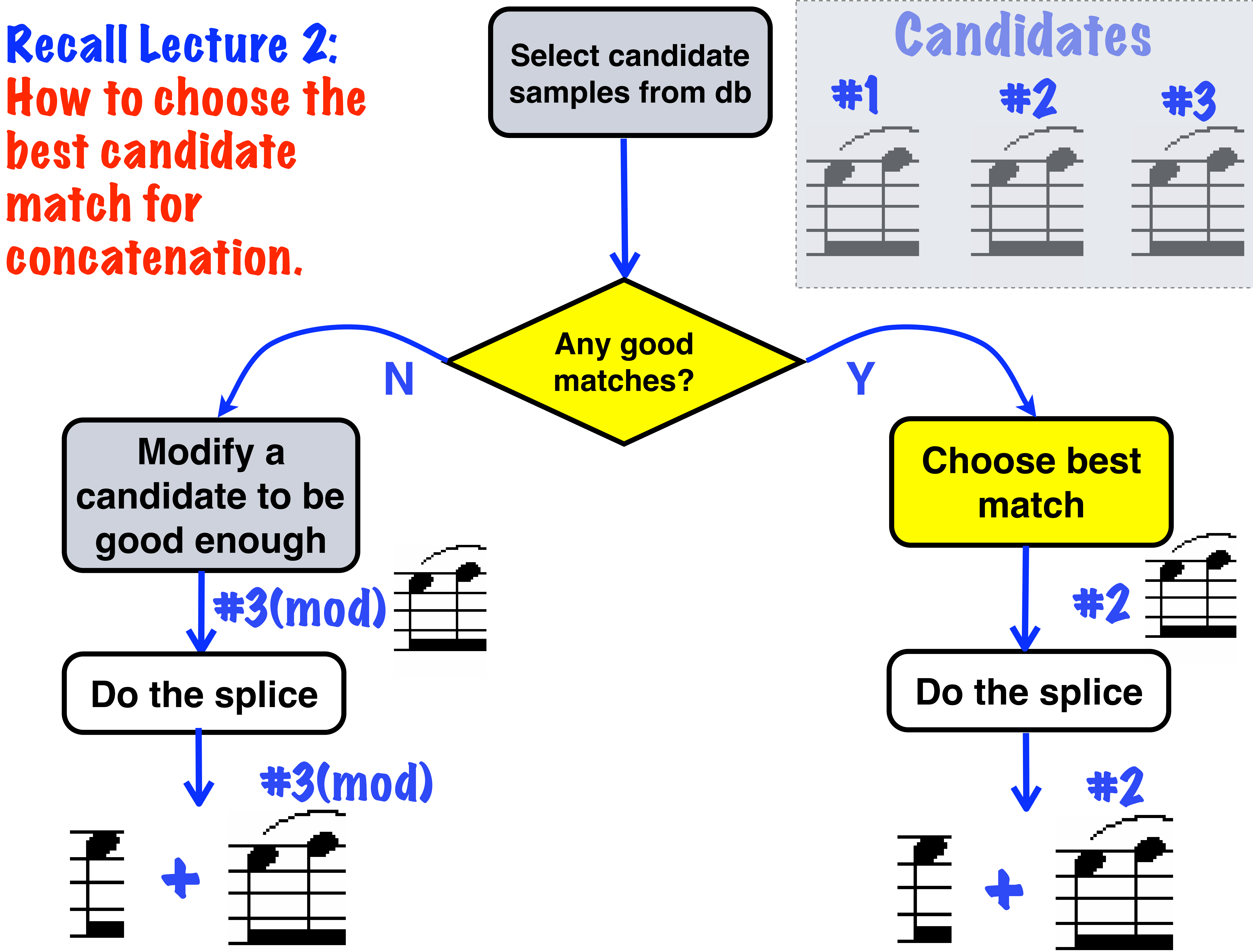
2006-3-2



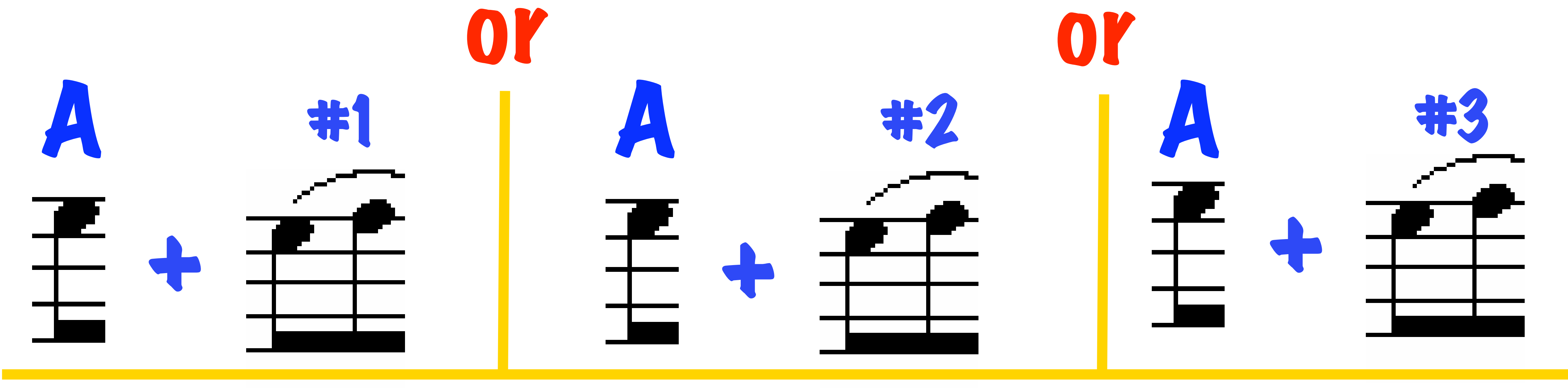
Professor David Wessel (with John Lazzaro)

(cnmat.berkeley.edu/~wessel, www.cs.berkeley.edu/~lazzaro)

`www.cs.berkeley.edu/~lazzaro/class/music209`


Recall Lecture 2:
How to choose the best candidate match for concatenation.
Select candidate samples from db
Candidates
#1
#2
#3
Any good matches?
N
Y
Modify a candidate to be good enough
#3(mod)
Do the splice
#3(mod)
+
Choose best match
#2
Do the splice
#2
+

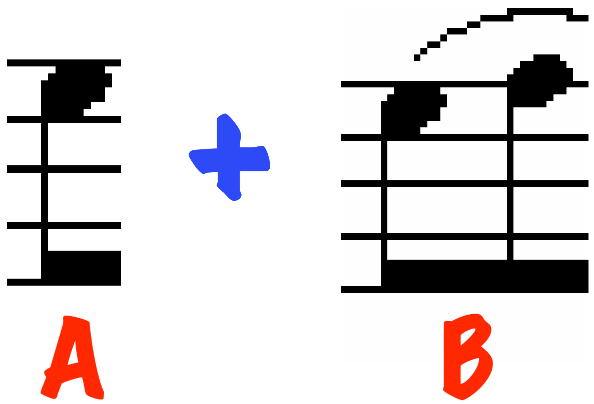
What makes a good (or the best) match?



Given samples A and B, we define a **metric f(A, B)** of **concatenation quality.** Compare f(A, #1), f(A, #2), f(A, #3) to find the best.

Compare best f() against an absolute standard to test for good enough.

What makes a transparent splice?



No waveform discontinuity at the splice point. Easy to handle in the "do the splice" algorithm.

Harder: The end of A and the start of B should have ...

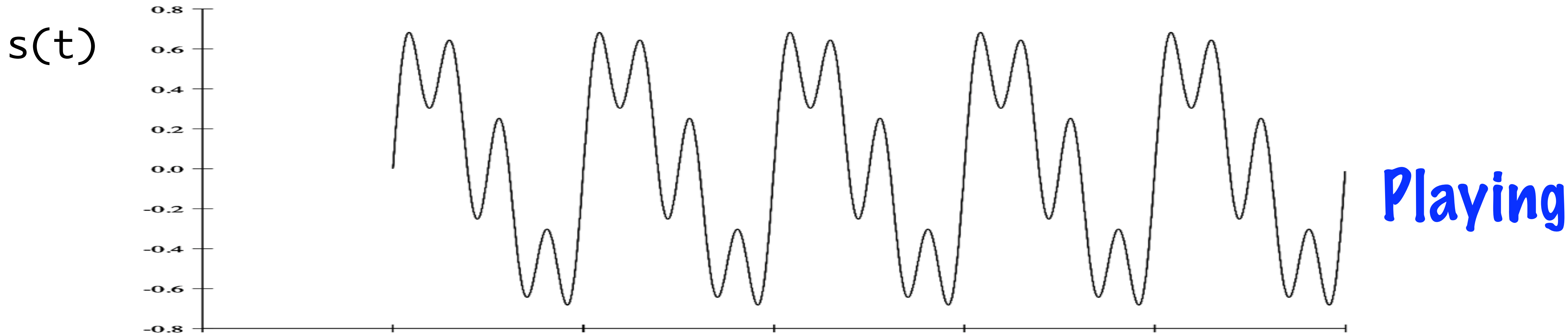
- **Similar loudness.**
  Absolute & delta: amplitude envelope, tremolo.
- **Similar spectral shape.**
  Absolute & delta: spectral motion across splice.
- **Similar pitch.**
  Absolute & delta: vibrato and pitch bends.

Topics for today ...

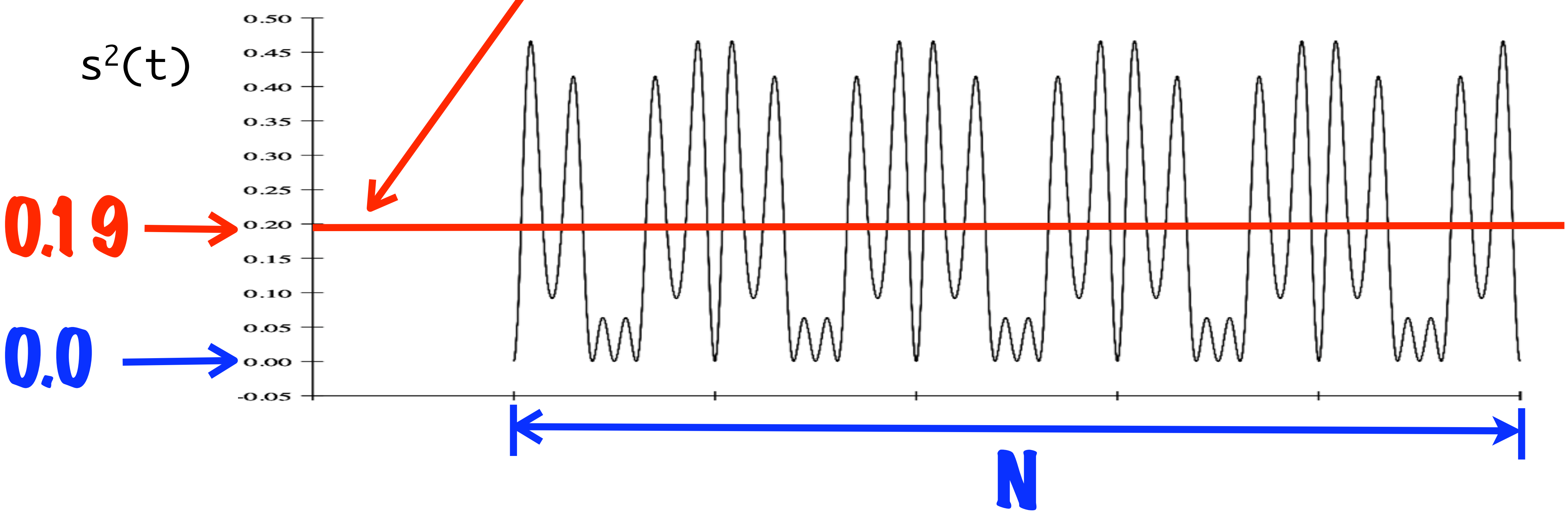
- **Database descriptors:** How to compute, access, and compare a waveform property for a database of audio waveform.
- **Energy** and **loudness** metrics.
- **Temporal structure** and **pitch** metrics.
- **Spectral shape** and **harmonic** metrics

Energy: RMS, dB's, and all that ...

s(t) = 0.5*sin(2*pi*t/1000) + 0.2*sin(2*pi*t/500) + 0.3*sin(2*pi*t/250)



$$Energy = \frac{1}{N}\sum_{i=1}^{N}(s_i)^2$$



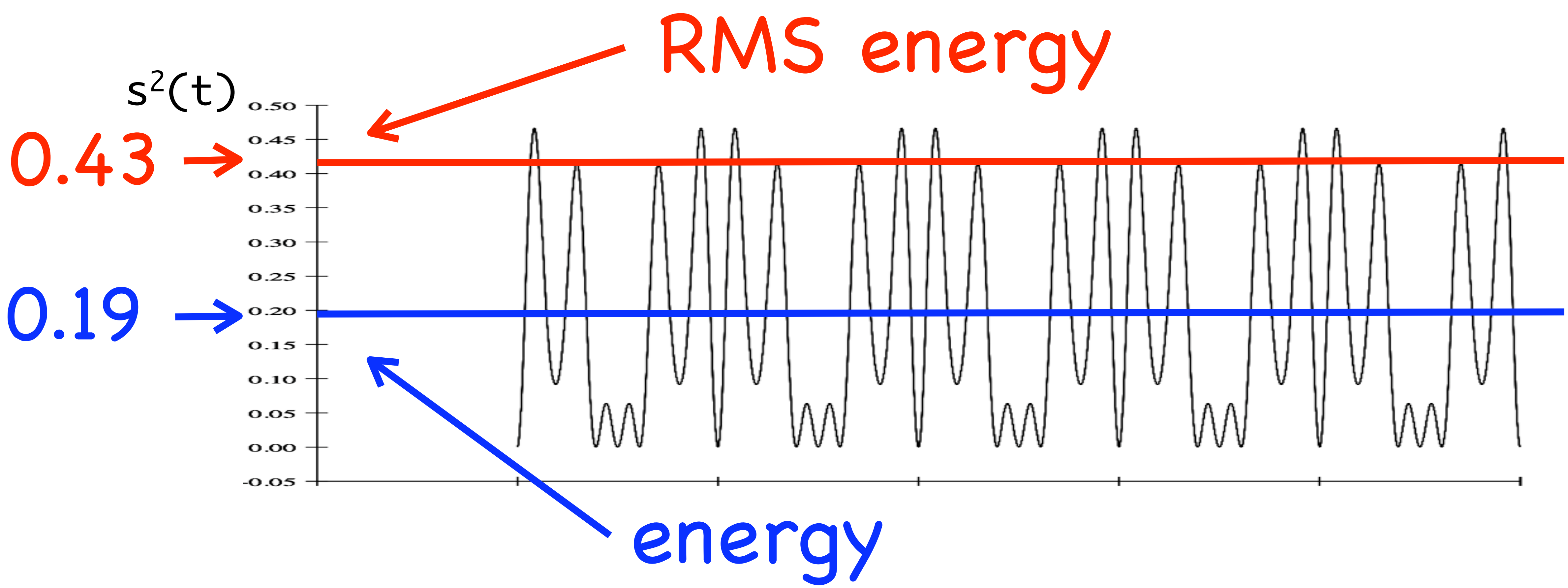
RMS energy ...

$$Energy = \frac{1}{N}\sum_{i=1}^{N}(s_i)^2$$

Units are amplitude^2

$$\text{RMS } Energy = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i)^2}$$

Units are amplitude



deciBels (dB) ...

Define (arbitrarily) 90dB as: $s_i(t) = 1.0$, for $\forall i$

Then ...

$$Energy = \frac{1}{N}\sum_{i=1}^{N}(s_i)^2$$

dB = 90 + 10 $\log_{10}$ (Energy)

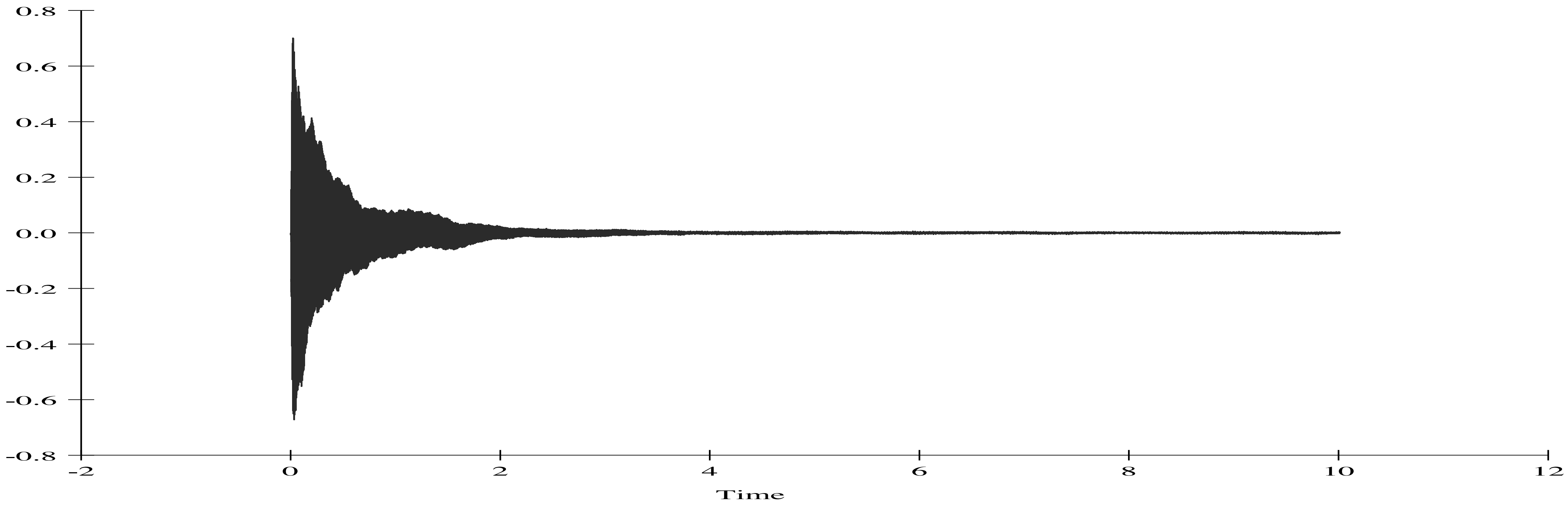
dB = 90 + 20 $\log_{10}$ (RMS E)

$s_i(t)$ for $\forall i$	Energy	$\log_{10}(E)$	dB
1.0	1.0	0	90
0.1	0.01	-2	70
0.01	0.0001	-4	50
0.001	0.000001	-6	30

Perception: Loudness of a sinusoid roughly goes as the cube root of amplitude -- close to logarithmic.

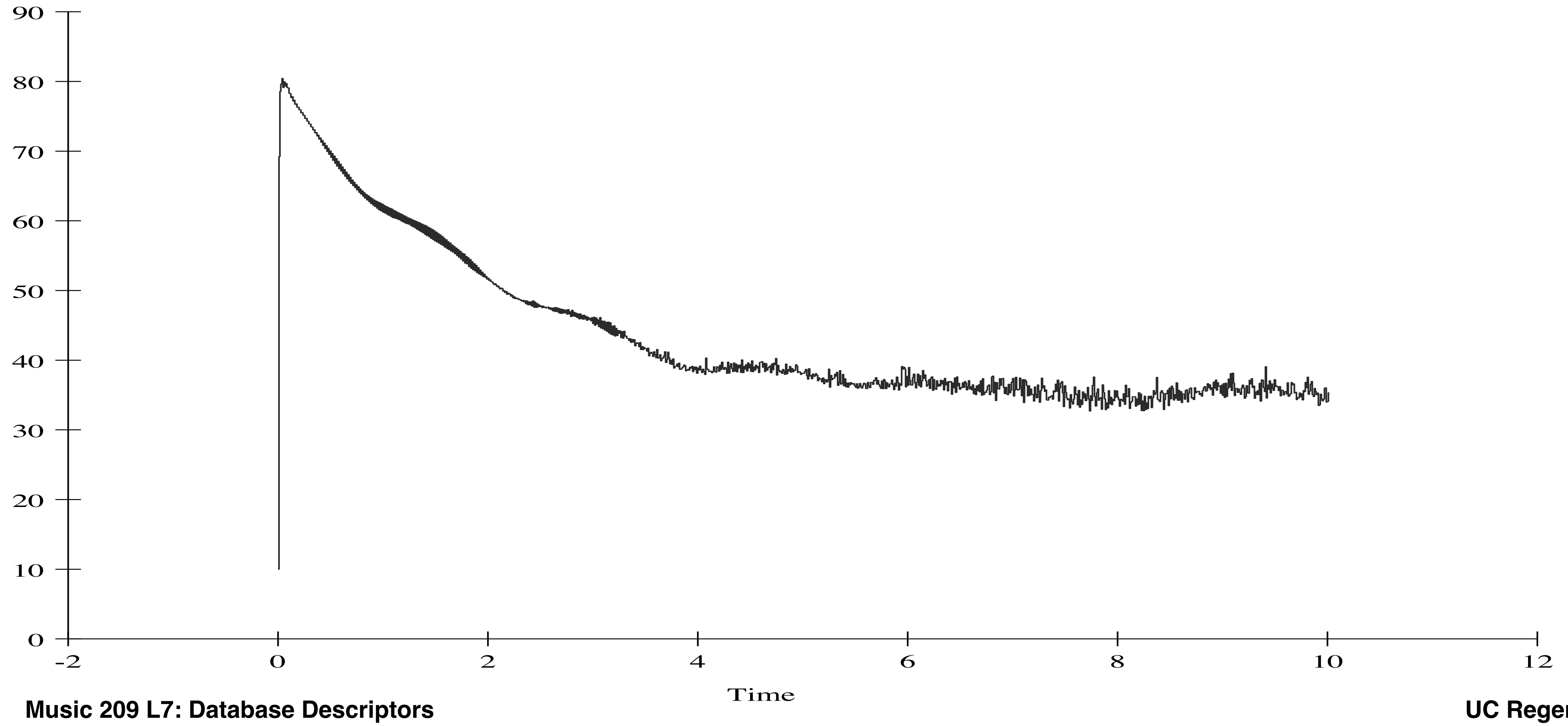
Example: GB Piano, Middle C, medium velocity

s(t)

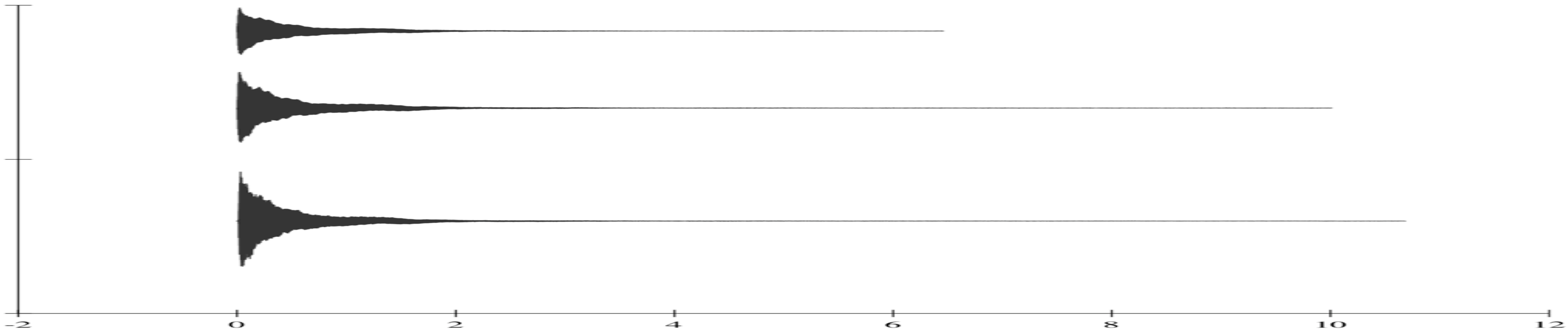


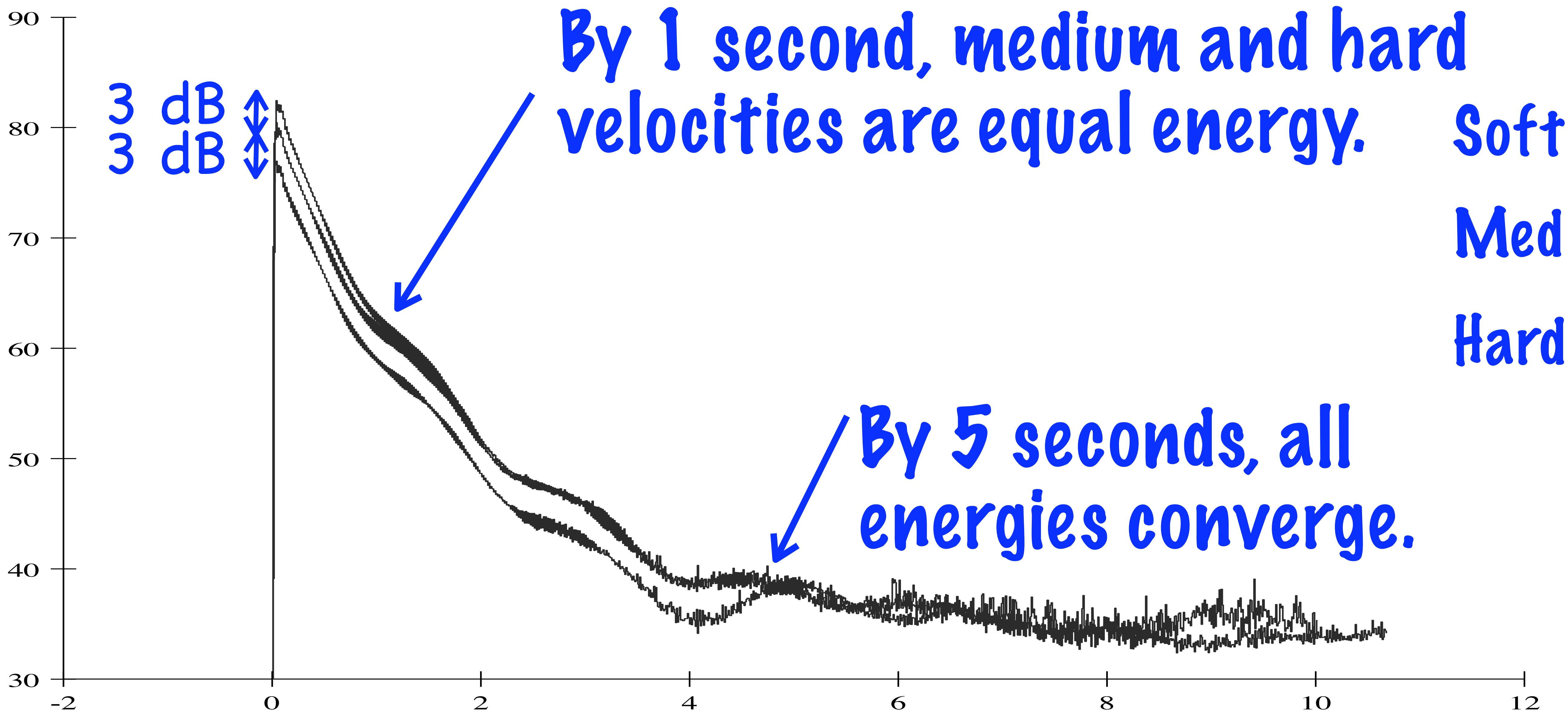
Playing

dB(t)

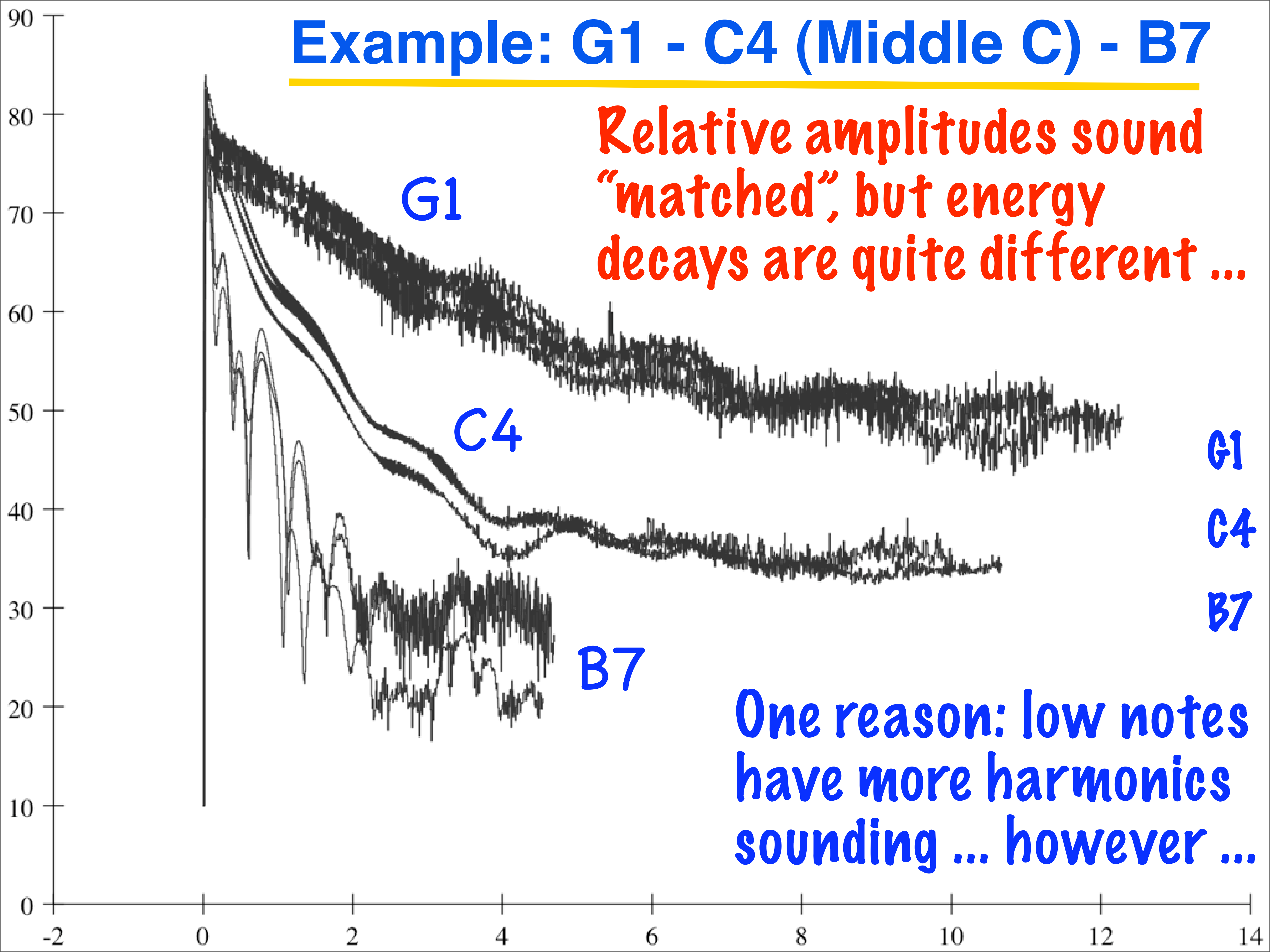


Example: Middle C, soft - medium - hard





Example: G1 - C4 (Middle C) - B7

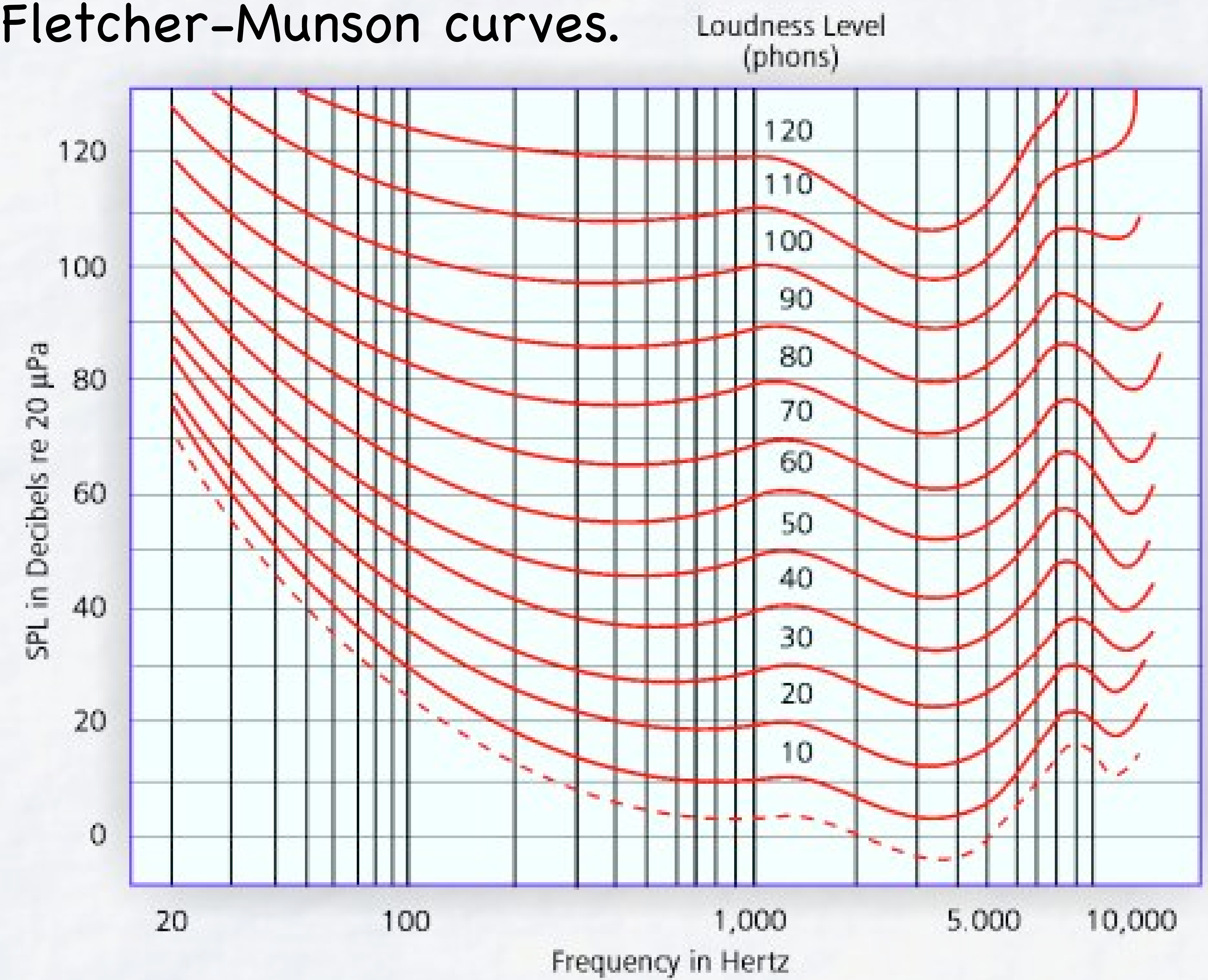


Relative amplitudes sound "matched", but energy decays are quite different ...

One reason: low notes have more harmonics sounding ... however ...

Loudness is also frequency-dependent …

Fletcher-Munson curves.

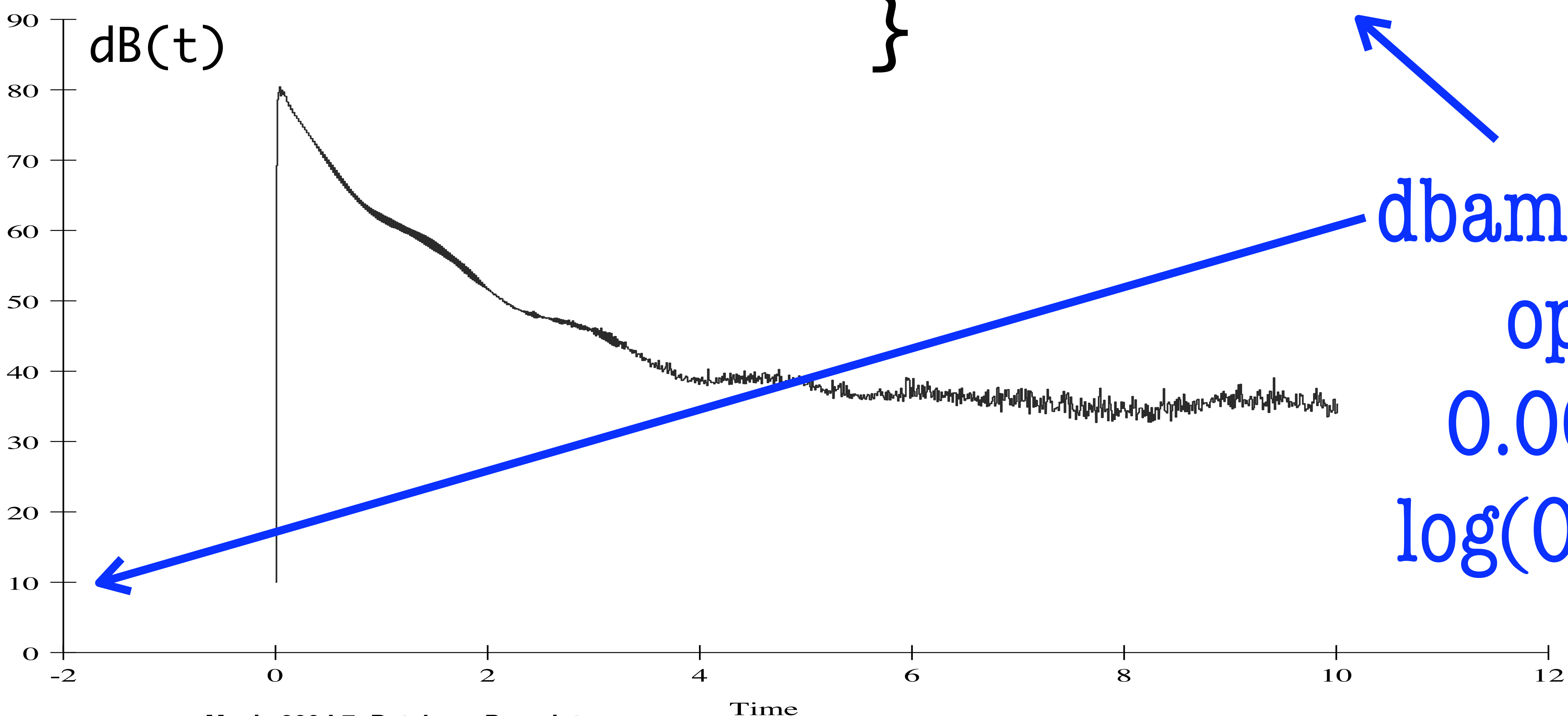


How we computed these graphs (in SAOL) ...

rms() is a "specialop". Accepts a-rate data, returns k-rate values.

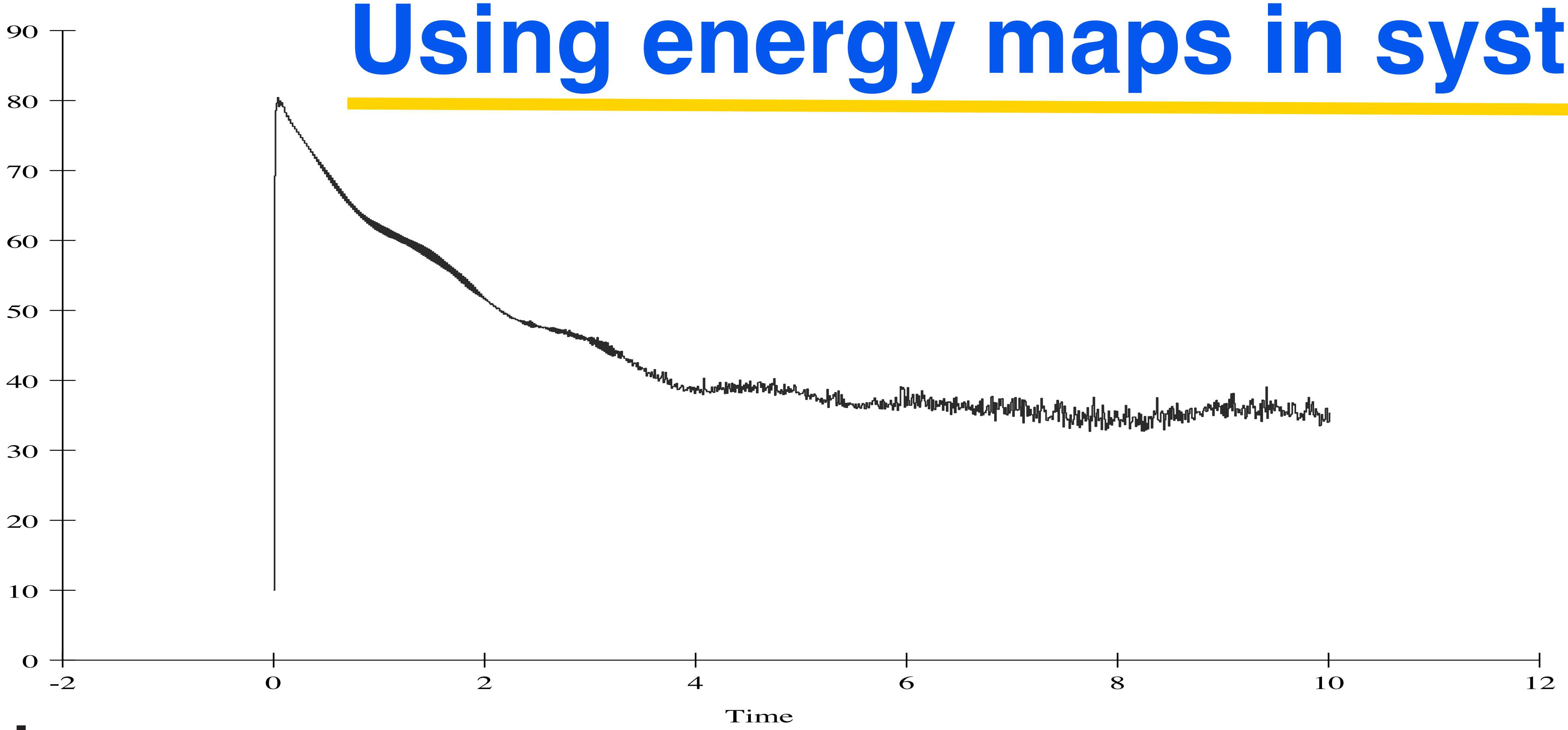
```
instr mixer () {
  ksig e, db;

  e = rms(input[0]);
  db = dbamp(e + 0.0001);
}
```



dbamp() is a core opcode for db. 0.0001 because log(0) undefined.

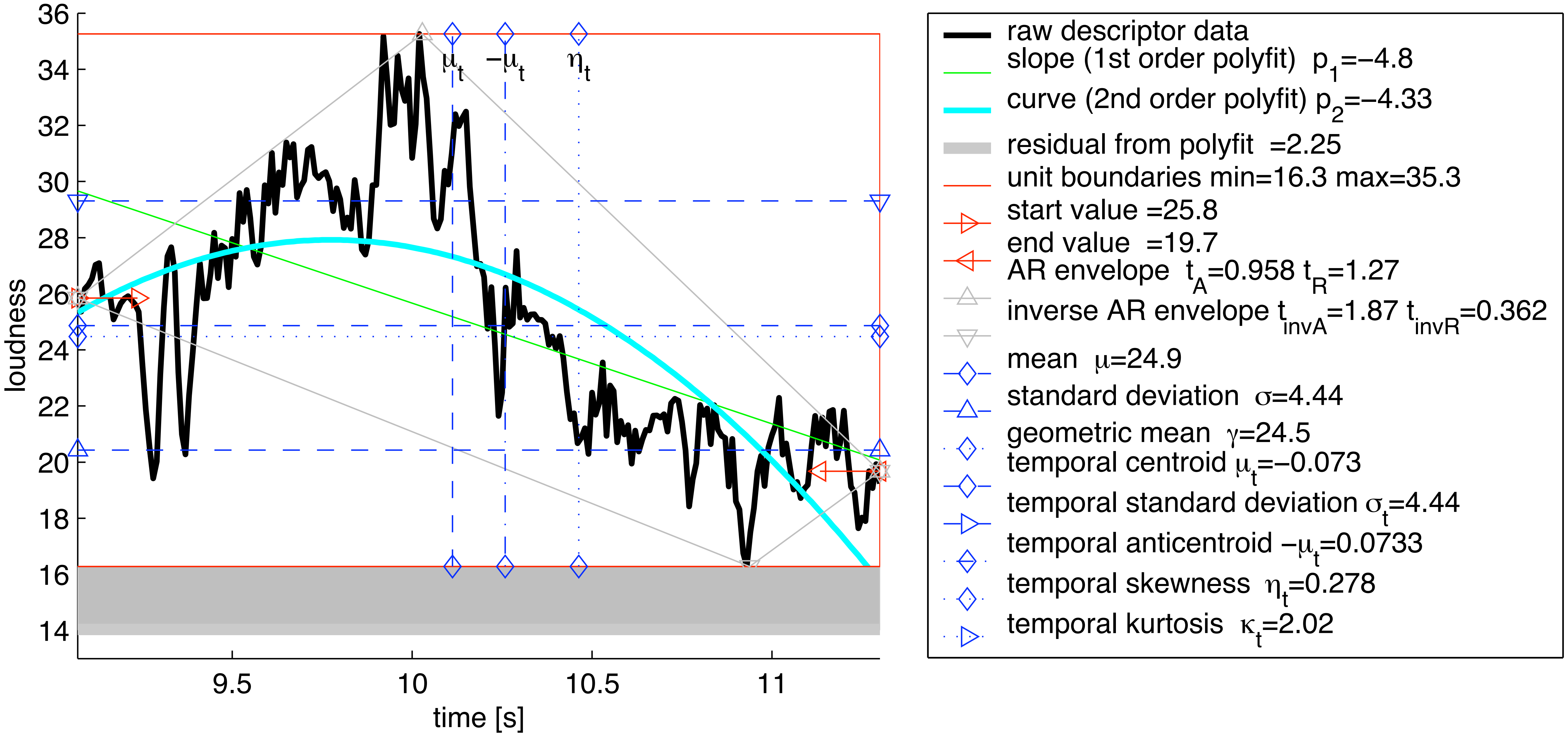
Using energy maps in systems



- **Shadow tables** of waveform (as above).
- **Inverse tables:** Table index is dB, table entry is a pointer to a waveform data point.
- **Delta features:** Time derivative of shadow dB table codes transients, steady sections.
- **Time filters:** 1-3 Hz, 3-5 Hz, 5-10Hz components.

"Modulation transfer functions" in neuroscience.

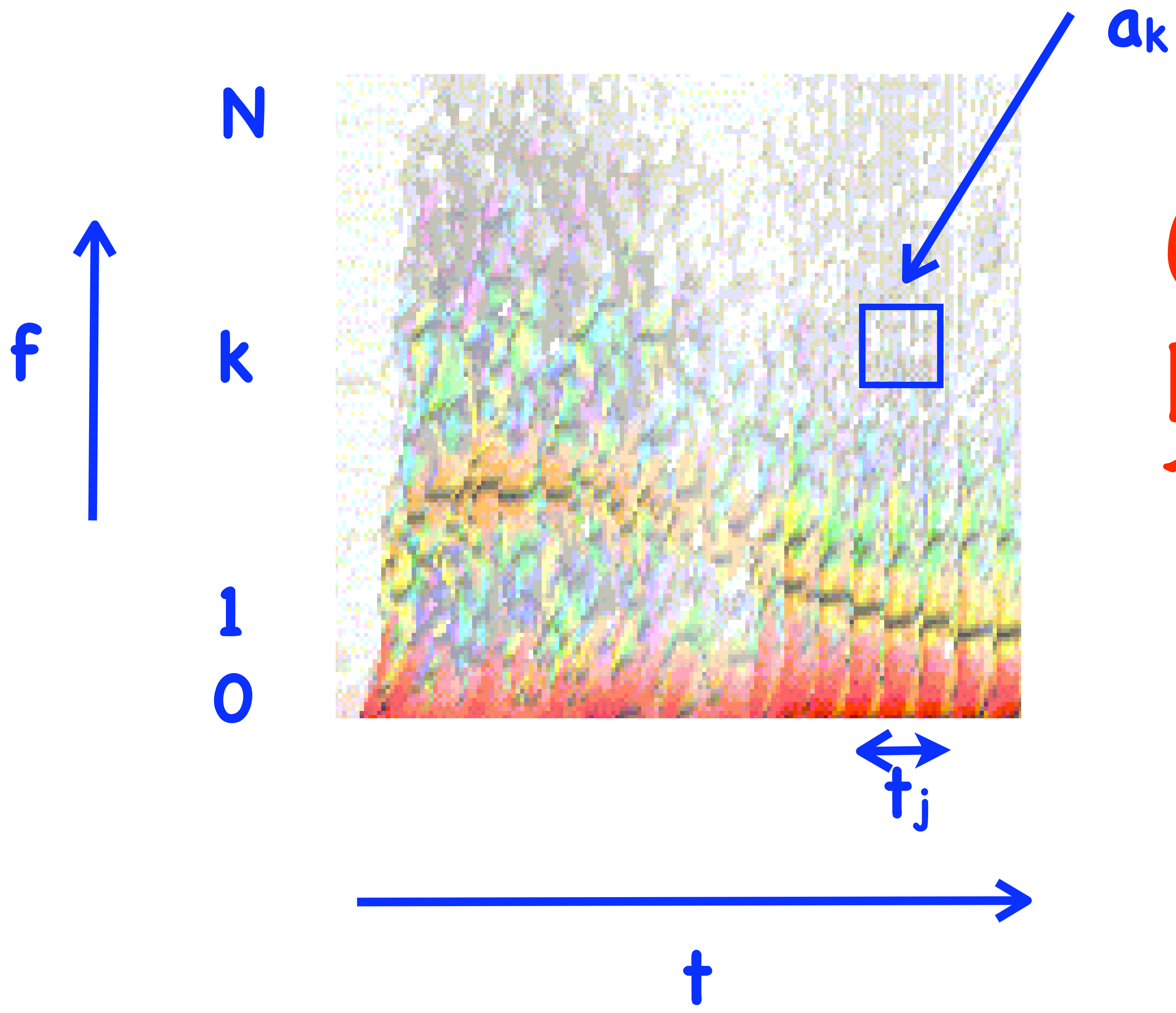
Reducing entire graph to a number ...



"Characteristic values": from Diemo Schwarz 's Ph.D.

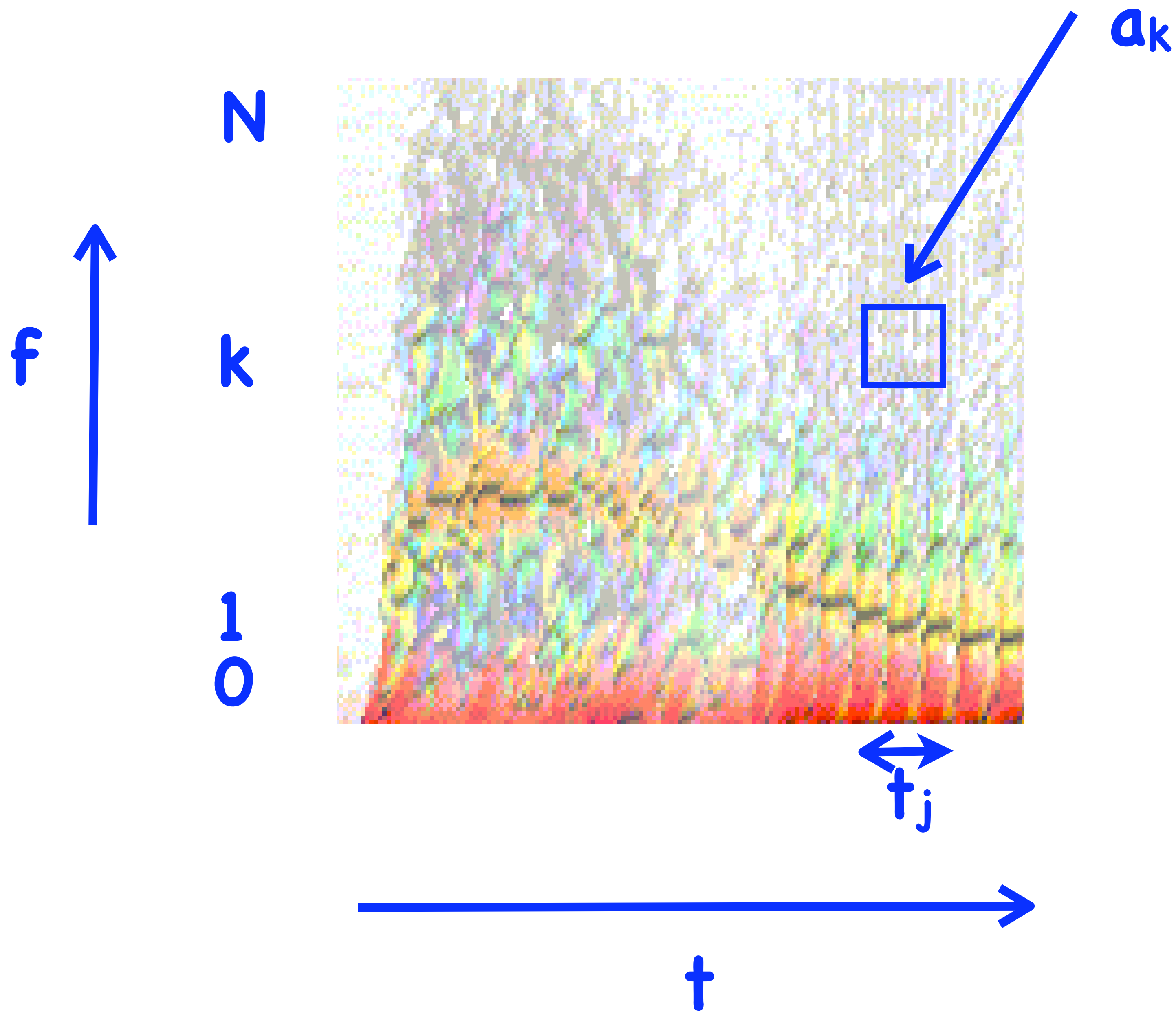
Spectral Shape

Summarizing spectrograms



Goal: Small set of parameters to describe the spectrum at time t_j.

Spectral Centroid



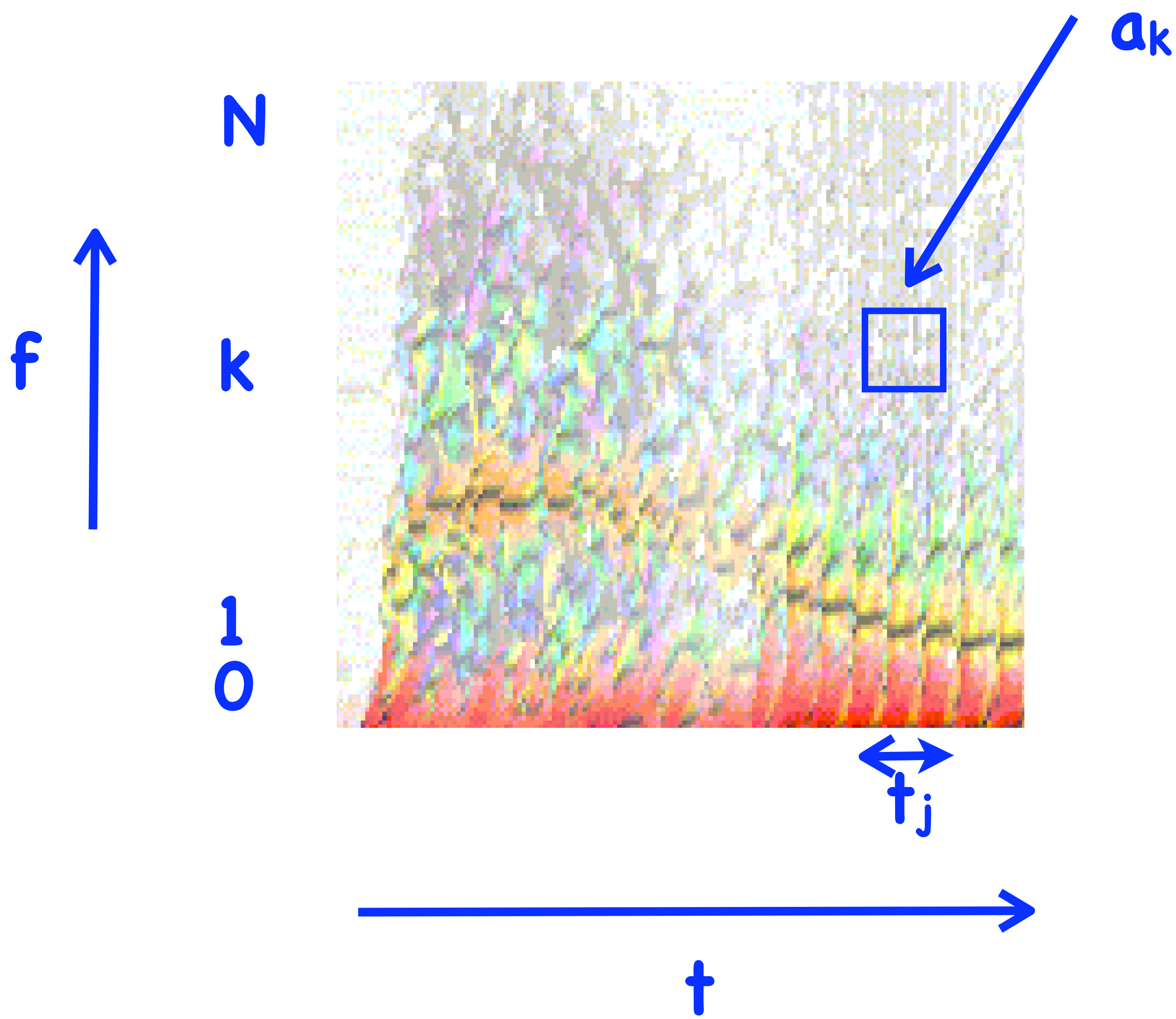
$$\frac{\sum_{i=1}^{N} a_i f_i}{\sum_{i=1}^{N} a_i}$$

Center of mass of the spectral slice. Related to the perception of brightness.

Also, "Harmonic centroid": computed on partials ...

Spectral Spread

$$\sqrt{\frac{\sum_{i=1}^{N} a_i \left(f_i - SpectralCentroid\right)^2}{\sum_{i=1}^{N} a_i}}$$

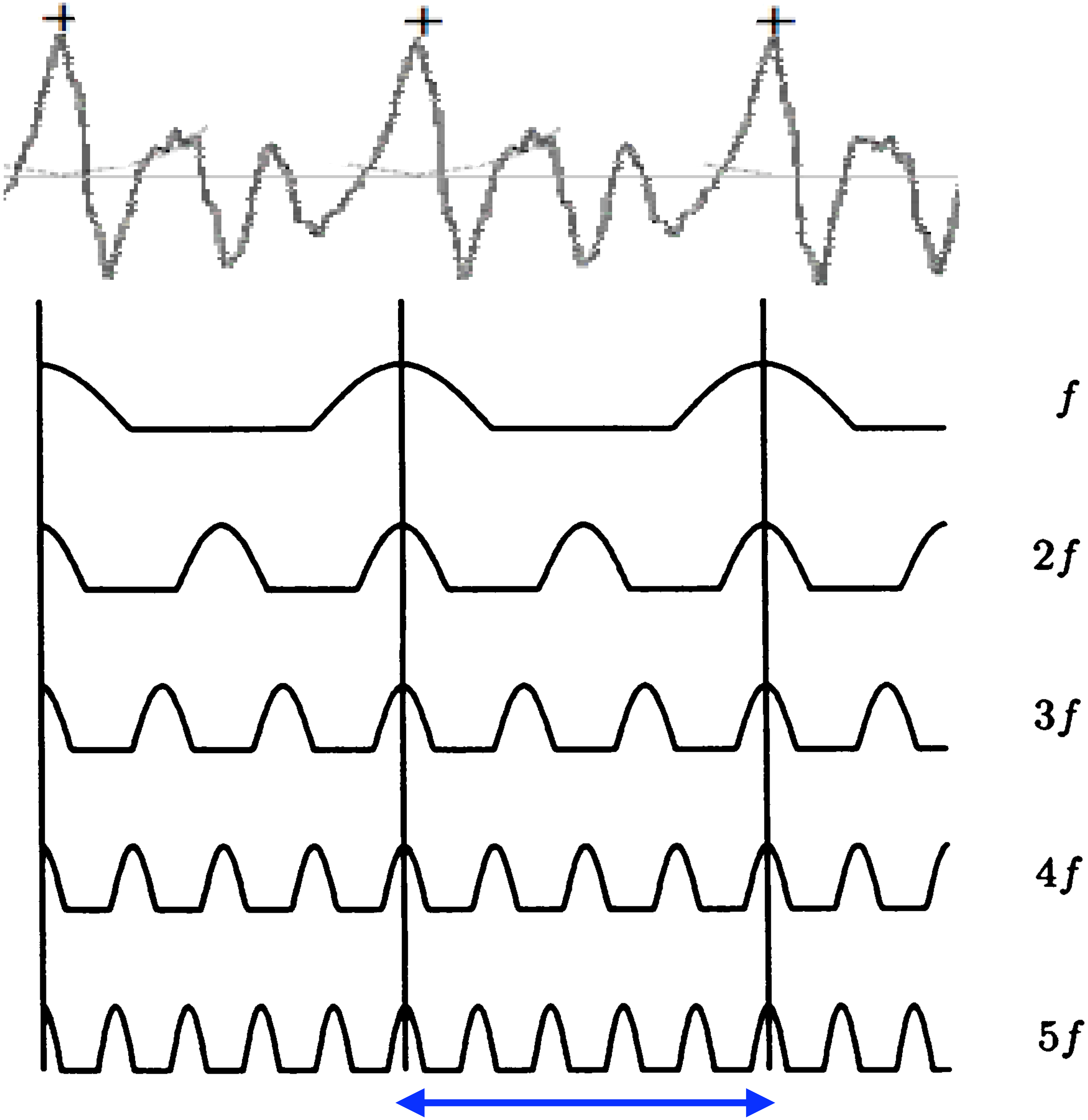


Standard deviation of spectral centroid.

See Diemo Schwarz 's Ph.D. for a complete list ...

Pitch

Summed waveform repeats at pitch frequency.

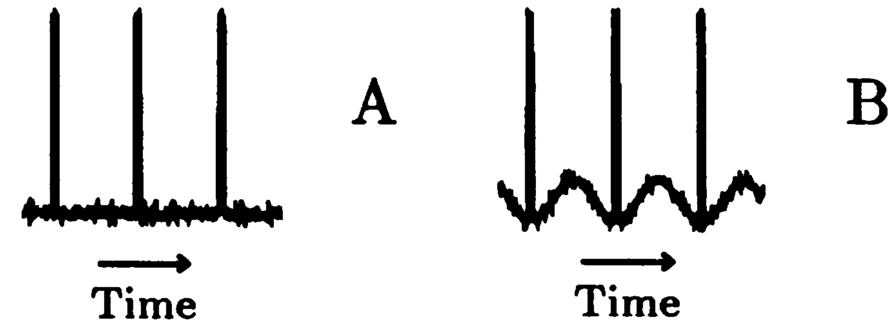


Frequencies of partials are integer multiples of an underlying fundamental.

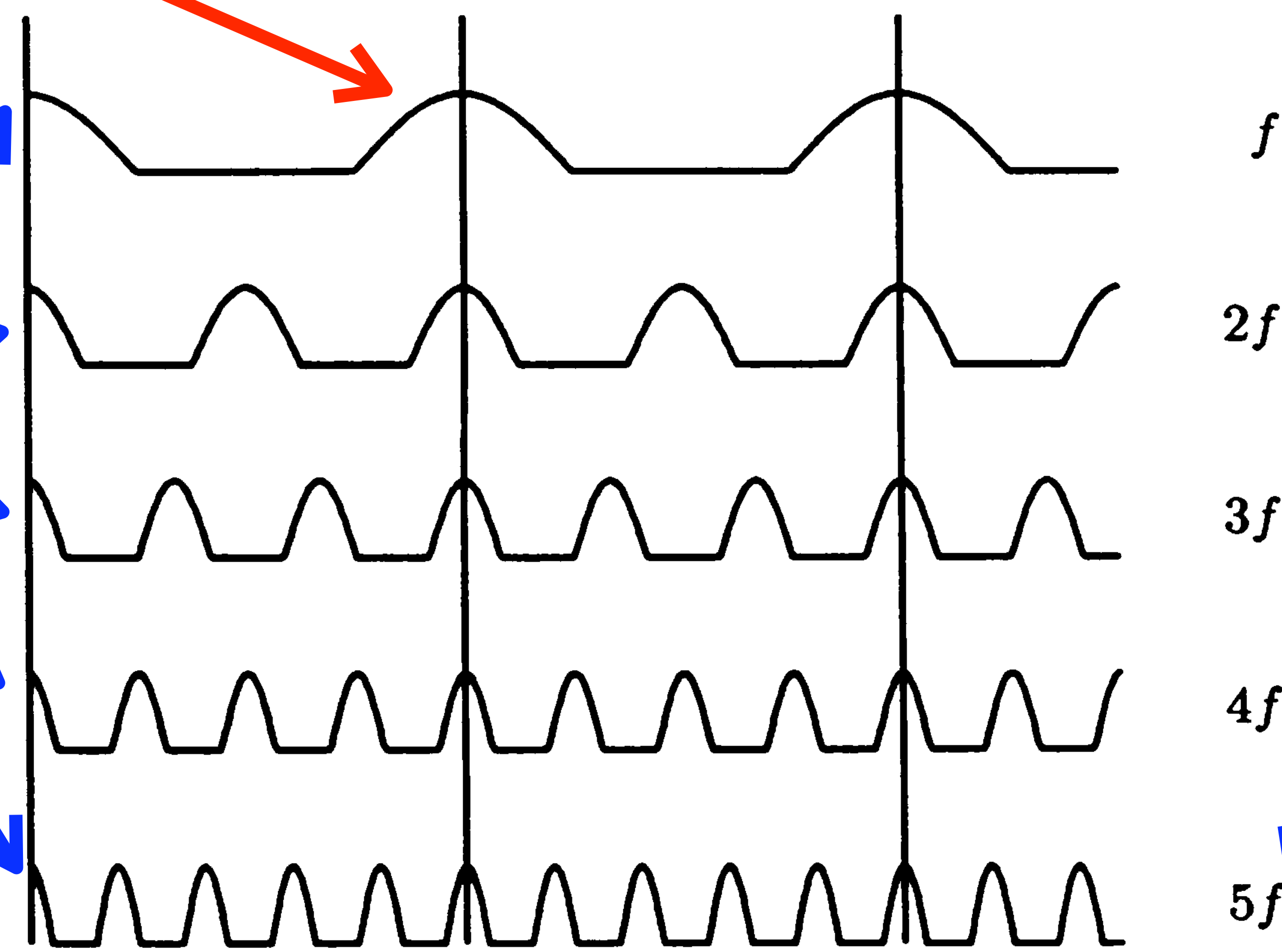
Pitch Period = 1/(Pitch Frequency)

Recall ...

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

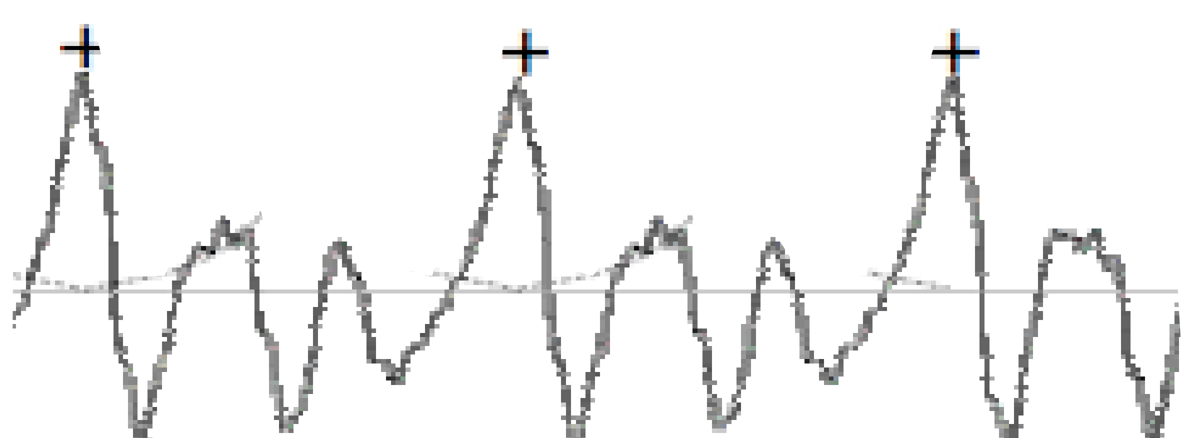


Relative phases of partials need not be aligned - any phase relation yields a strong pitch.

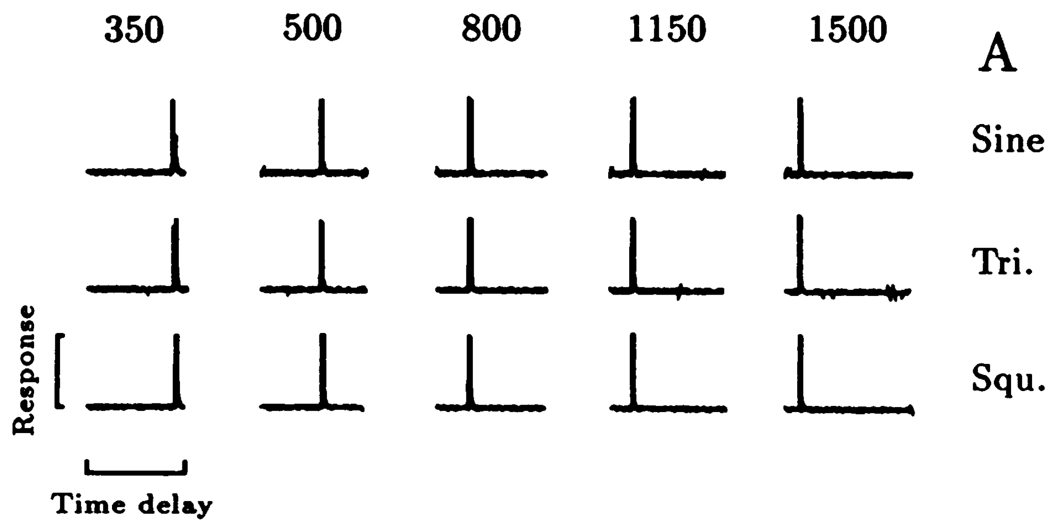


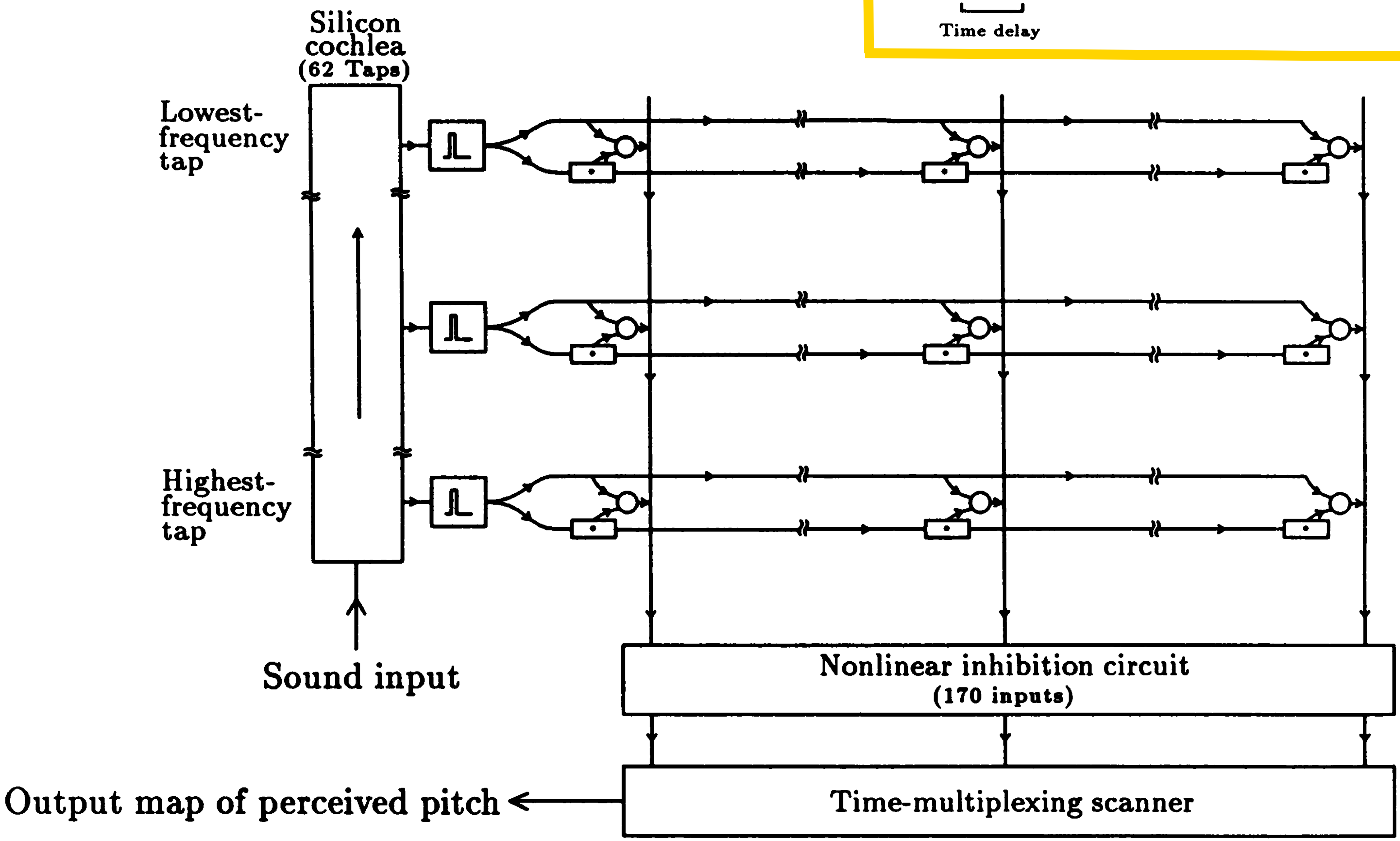
Sounds whose partials are not quite integer - related still yield a sense of pitch -

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



Computing pitch

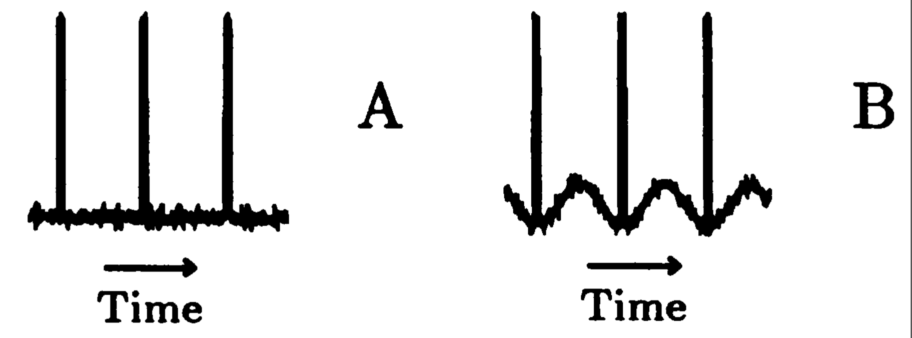




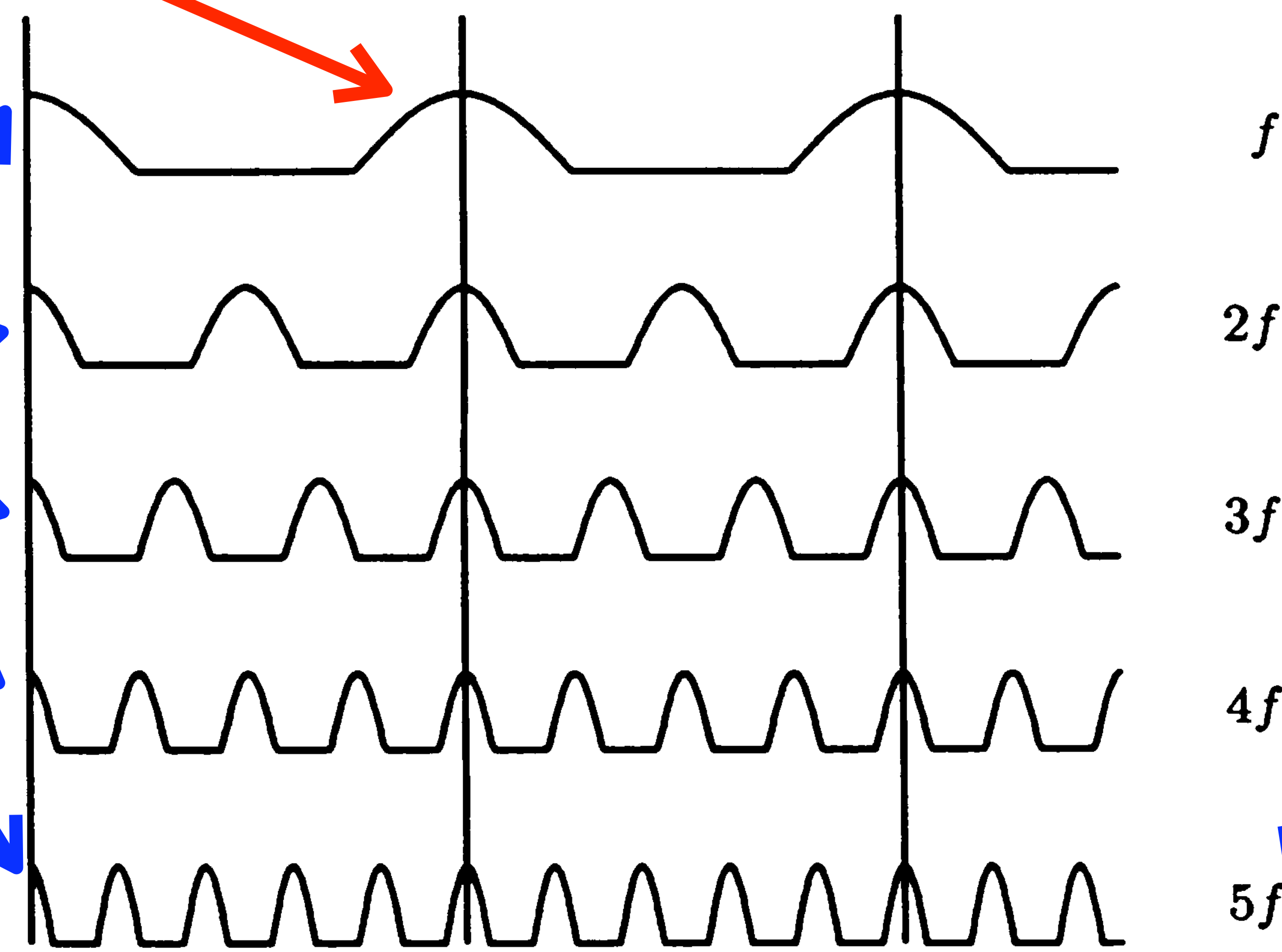
Licklider model: Autocorrelate filtered waveforms.

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 -

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