AUDL 4007
Auditory Perception

Week 2

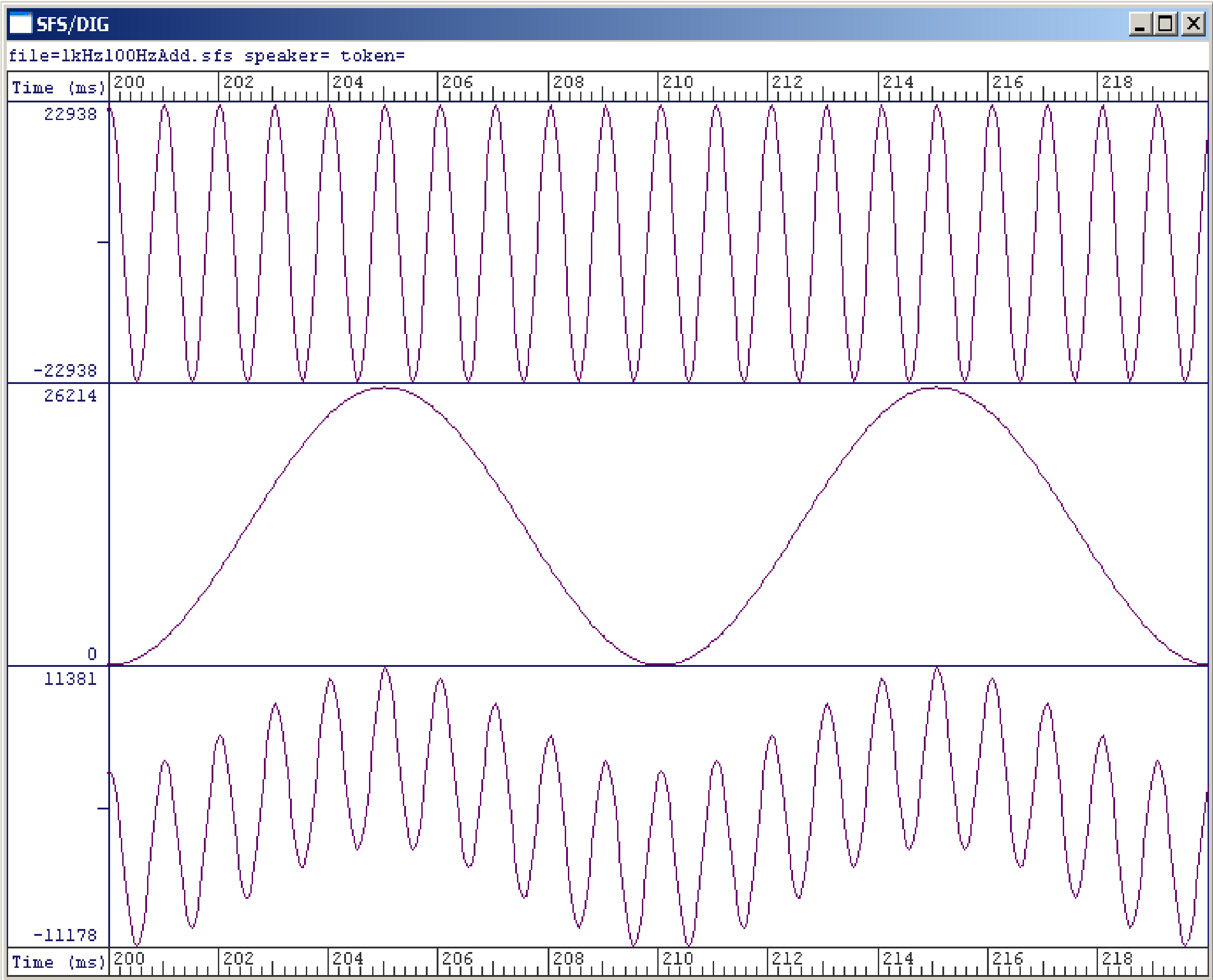
Envelope and temporal fine structure (TFS)

Envelope and TFS arise from a method of decomposing waveforms

The ‘classic’ decomposition of waveforms

- Spectral analysis …
  - Decomposes a complex wave into a sum of sinusoids to give a *spectrum*

Adding waves (time domain)



1 kHz sinusoid

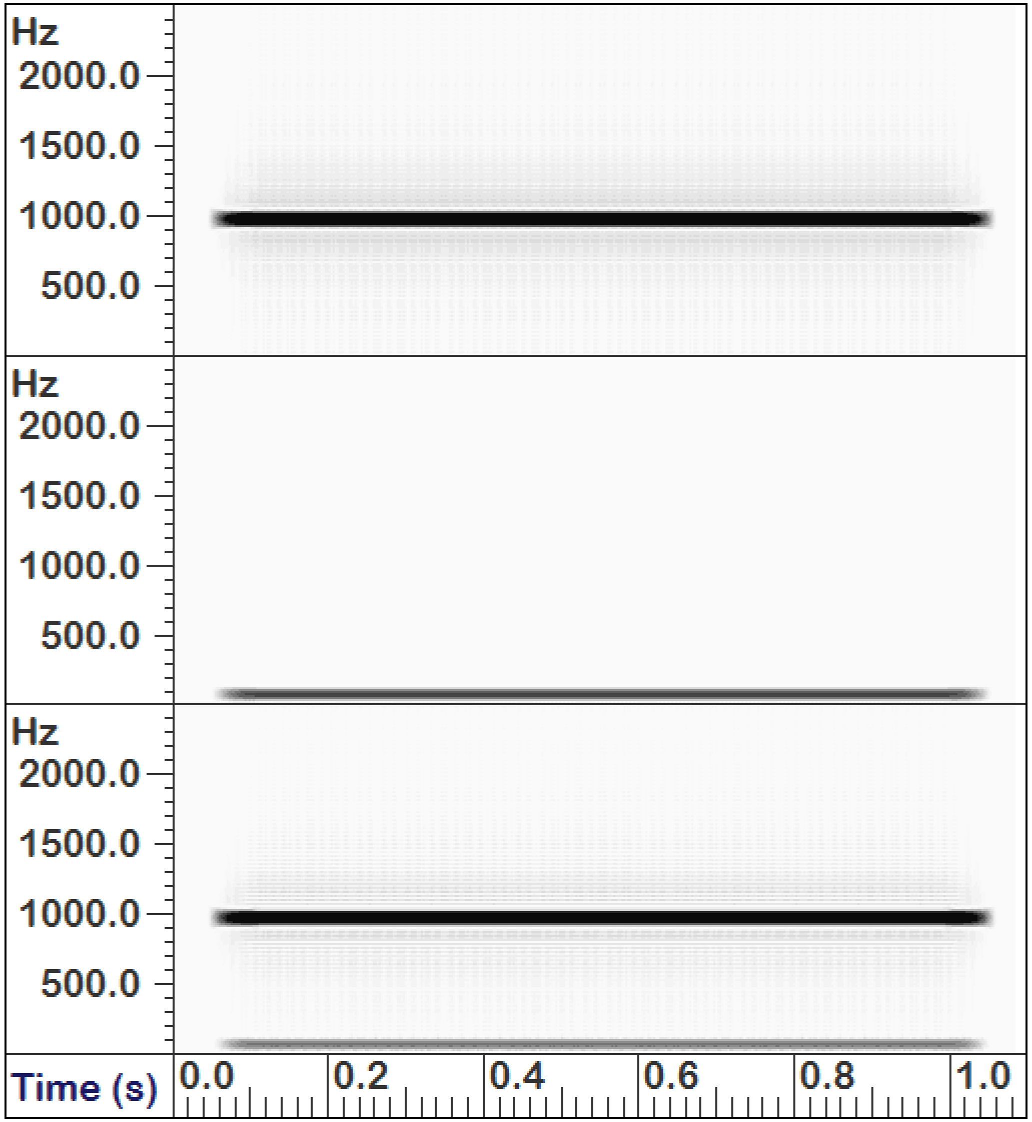
+

100 Hz sinusoid

=

a complex wave
(with two spectral
components)

Adding waves (frequency domain)



1 kHz sinusoid

+

100 Hz sinusoid

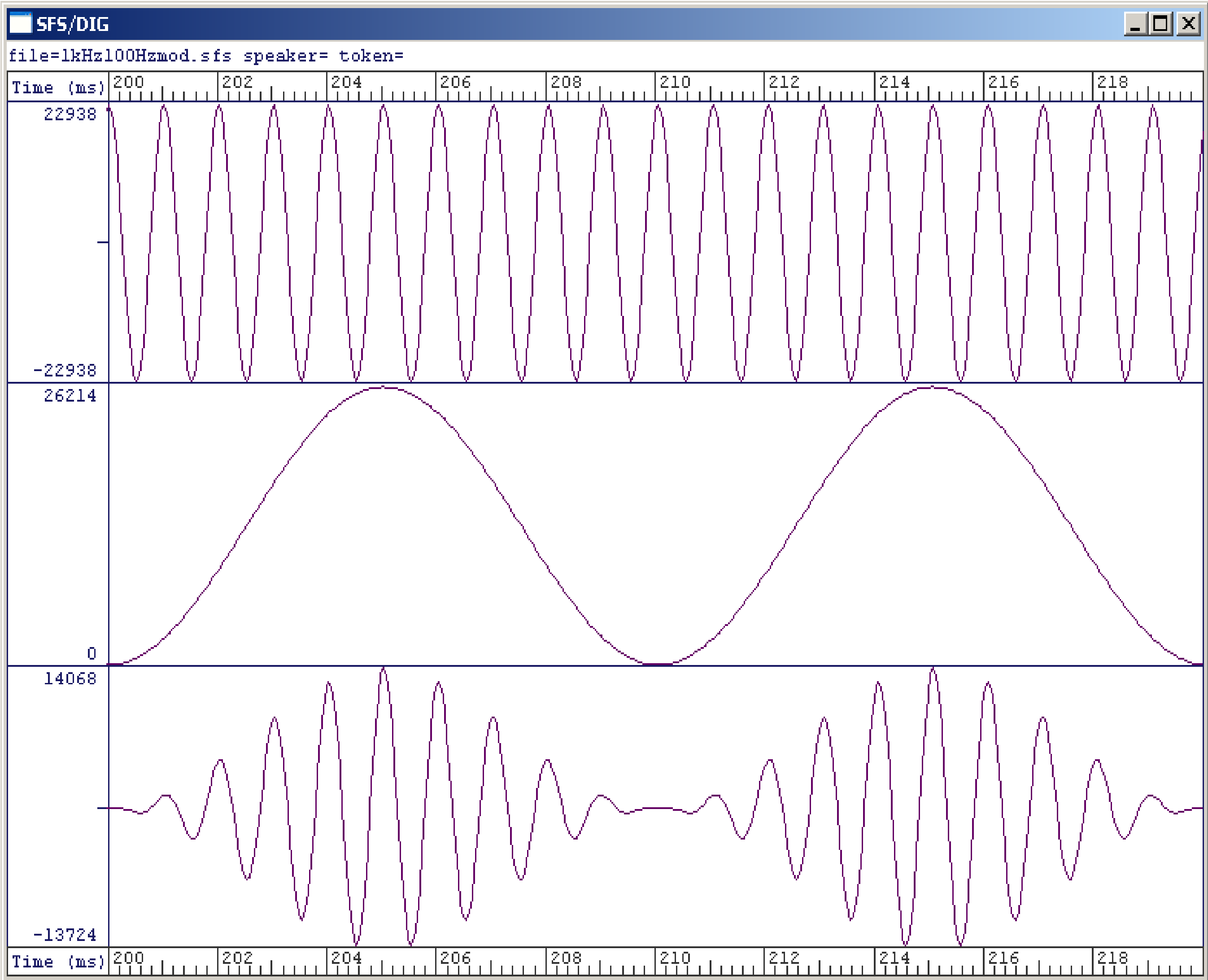
=

a complex wave
(with two spectral
components)

A less familiar way of decomposing waveforms in the time domain …

based on *multiplication*.

Multiplying (*modulating*) waves



carrier at 1 kHz
(fine structure)

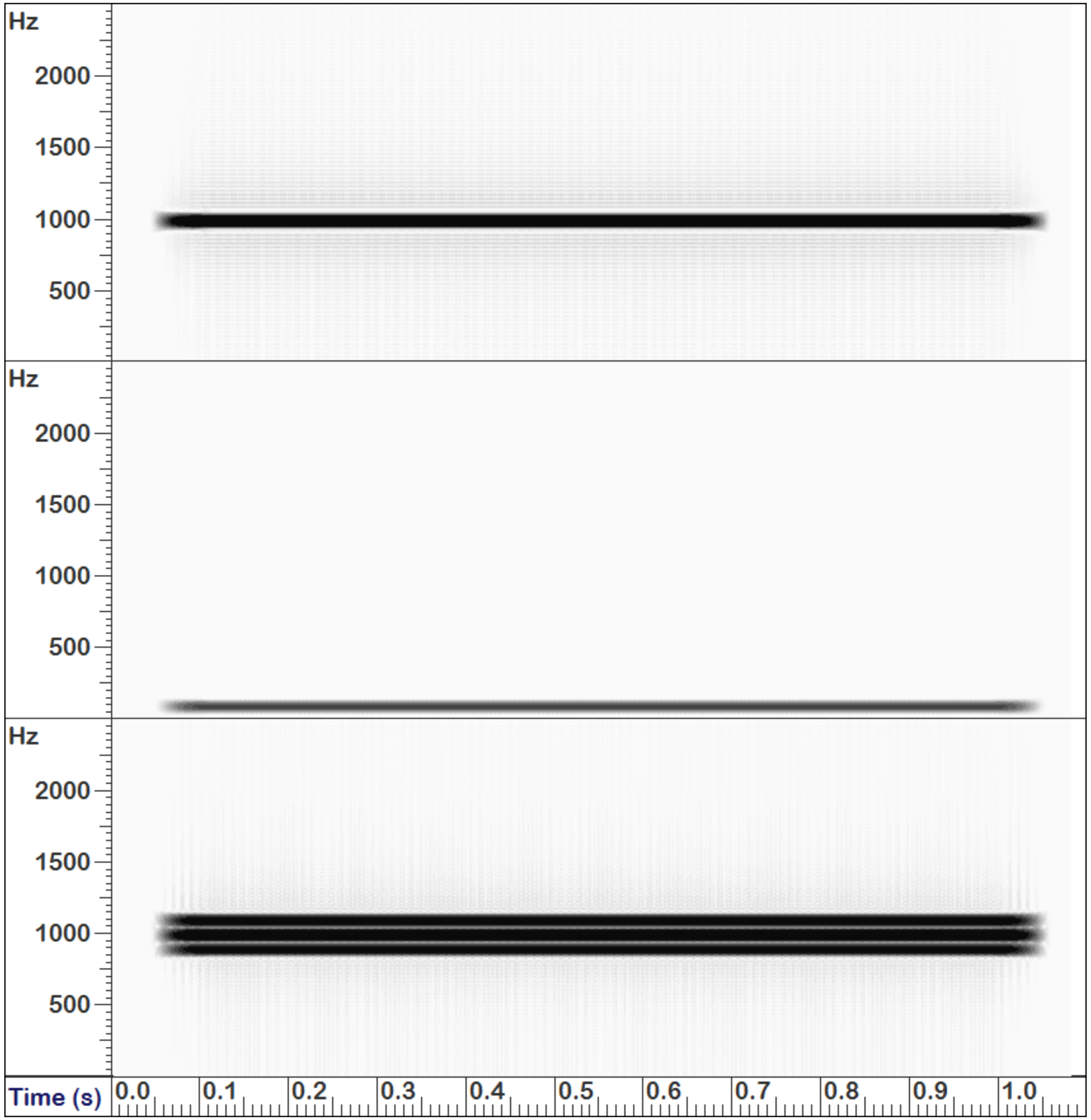
x

modulator at 100 Hz
(envelope)

=

amplitude-modulated wave

Multiplying (*modulating*) waves



carrier at 1 kHz
(fine structure)

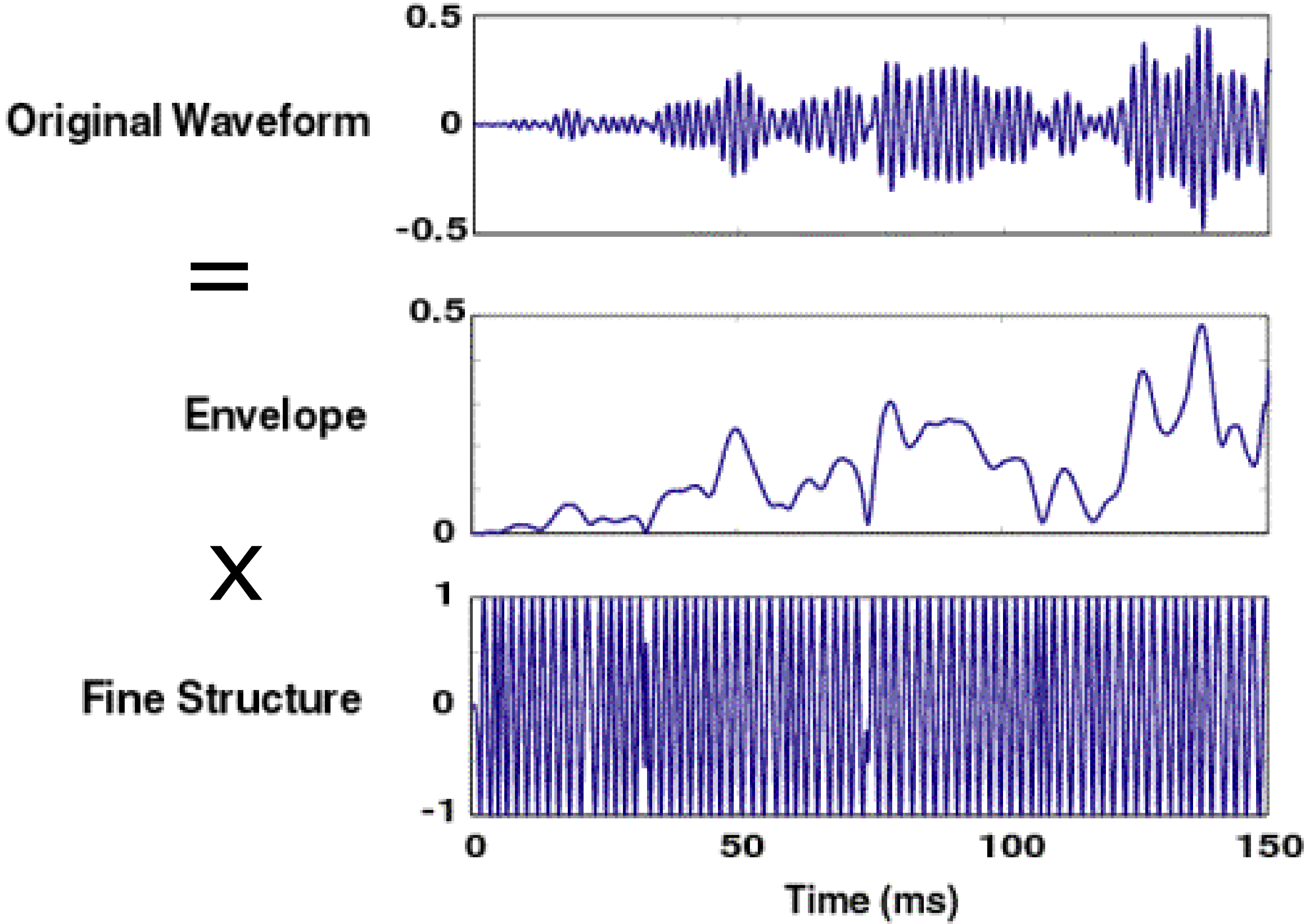
x

modulator at 100 Hz
(envelope)

=

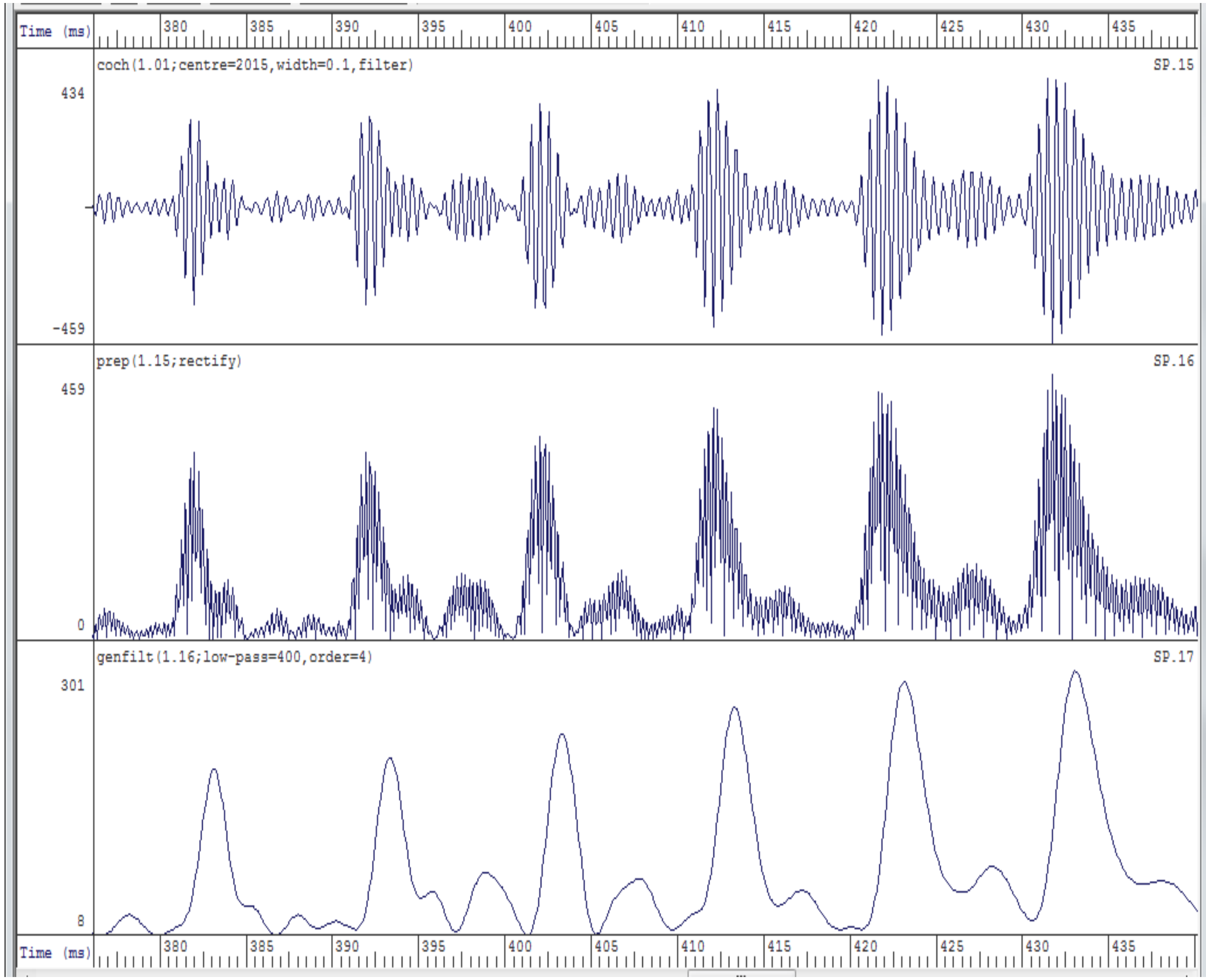
amplitude-modulated wave

Can work this backwards too



http://research.meei.harvard.edu/Chimera/motivation.html 24 JAN 2010

Extracting envelopes



original wave

full-wave rectification

smoothing at 400 Hz
(low-pass filtering)

A Hilbert transform

- can uniquely decompose a wave into the *product* of two waves
  - *envelope*
  - *temporal fine structure* (TFS)
- Unlike spectral analysis, the constituent waves are usually complicated
- A warning!

The outcome of a Hilbert decomposition

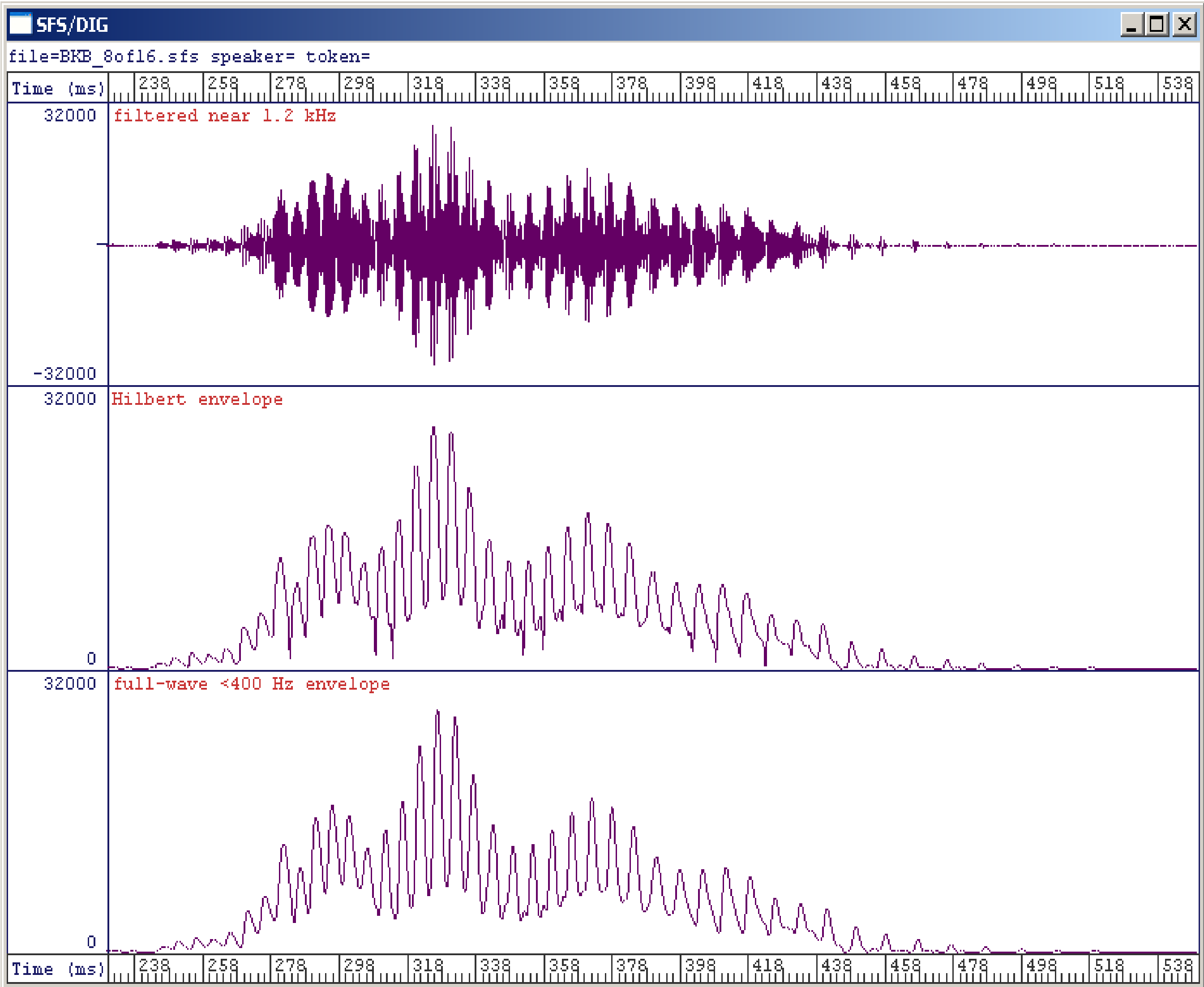
$$x(t) = ENV(t) \bullet \sin[2\pi f t + \Theta(t)]$$

a time-varying envelope

a constant amplitude sinusoid varying in frequency/phase

think of all waves as being made by multiplying one wave (the *envelope*) against another (the *temporal fine structure*)

There's more than one way to extract an envelope

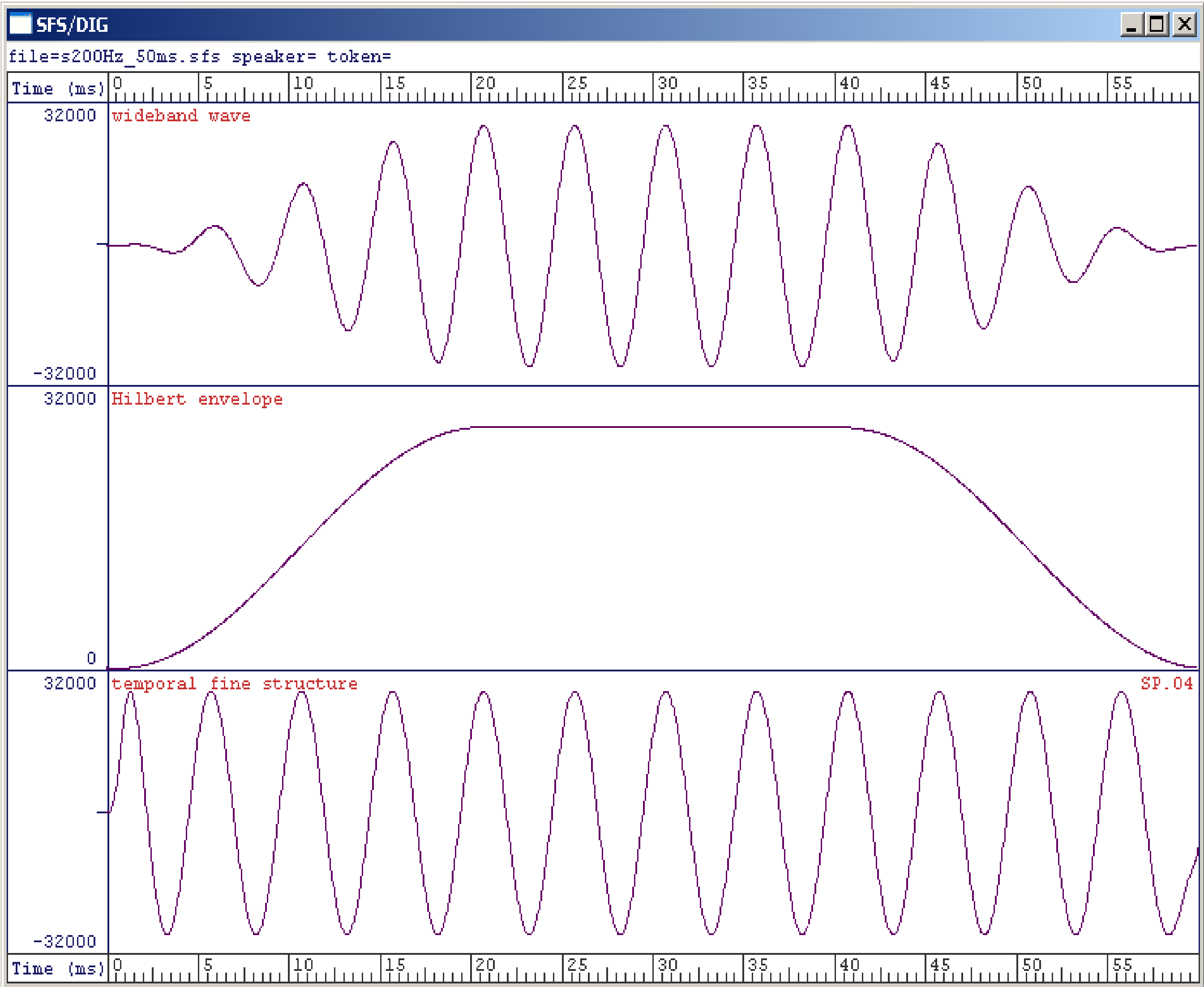


original wave

Hilbert envelope

envelope from full-wave rectification and smoothing at 400 Hz

A simple example: a tone pulse



original wave

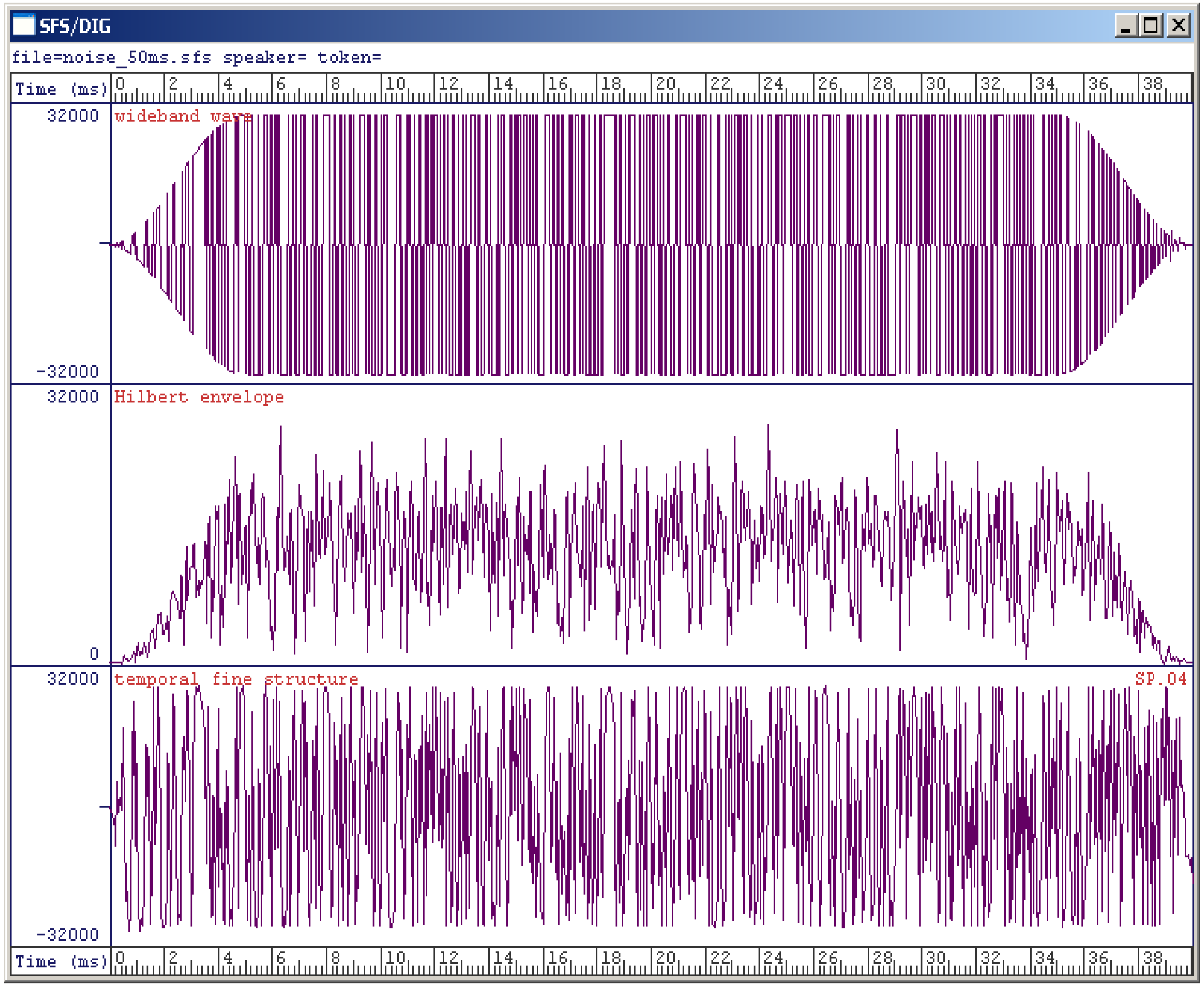
=

envelope

x

TFS

A simple example: a noise pulse



original wave

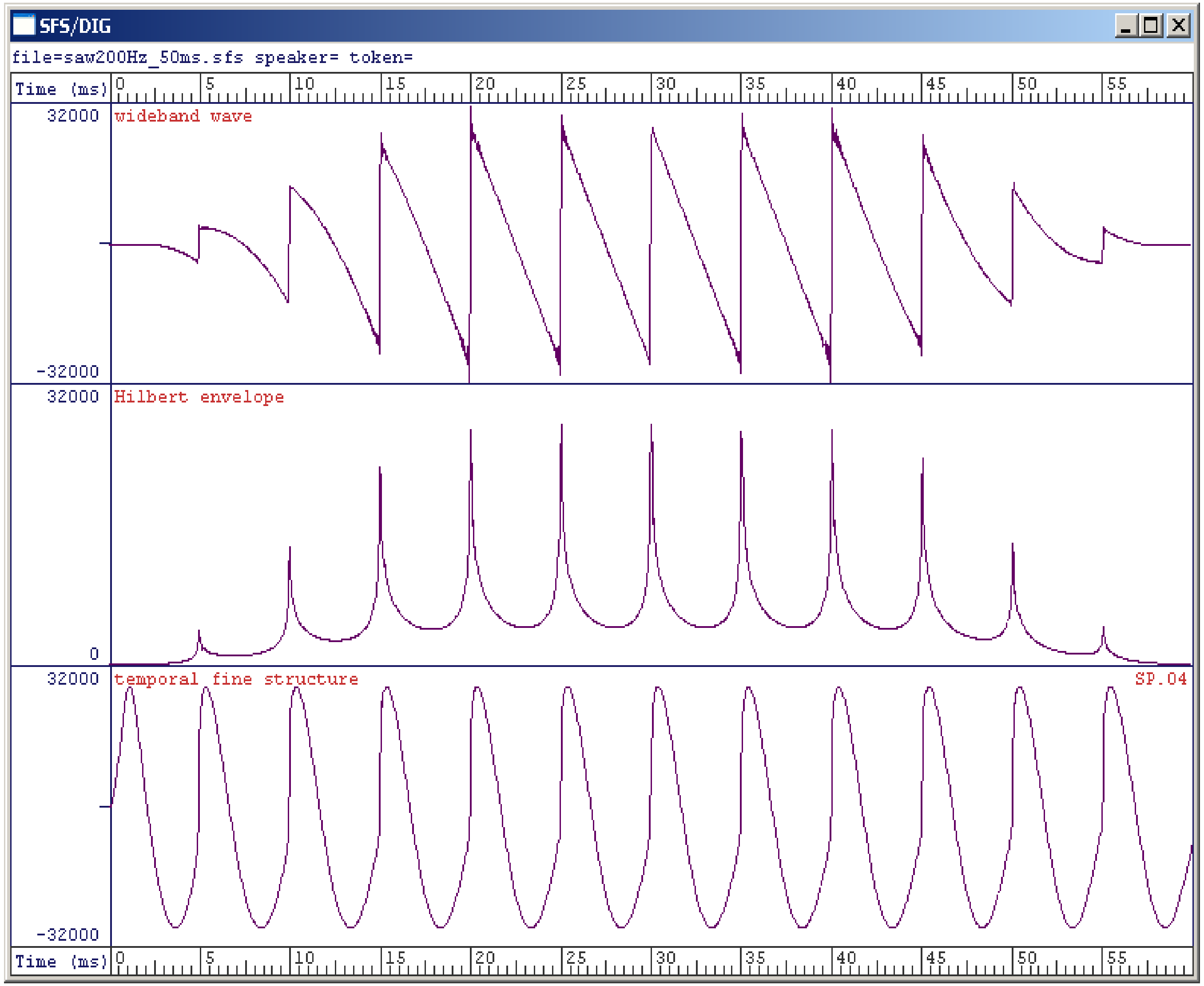
=

envelope

x

TFS

A simple example: a sawtooth



original wave

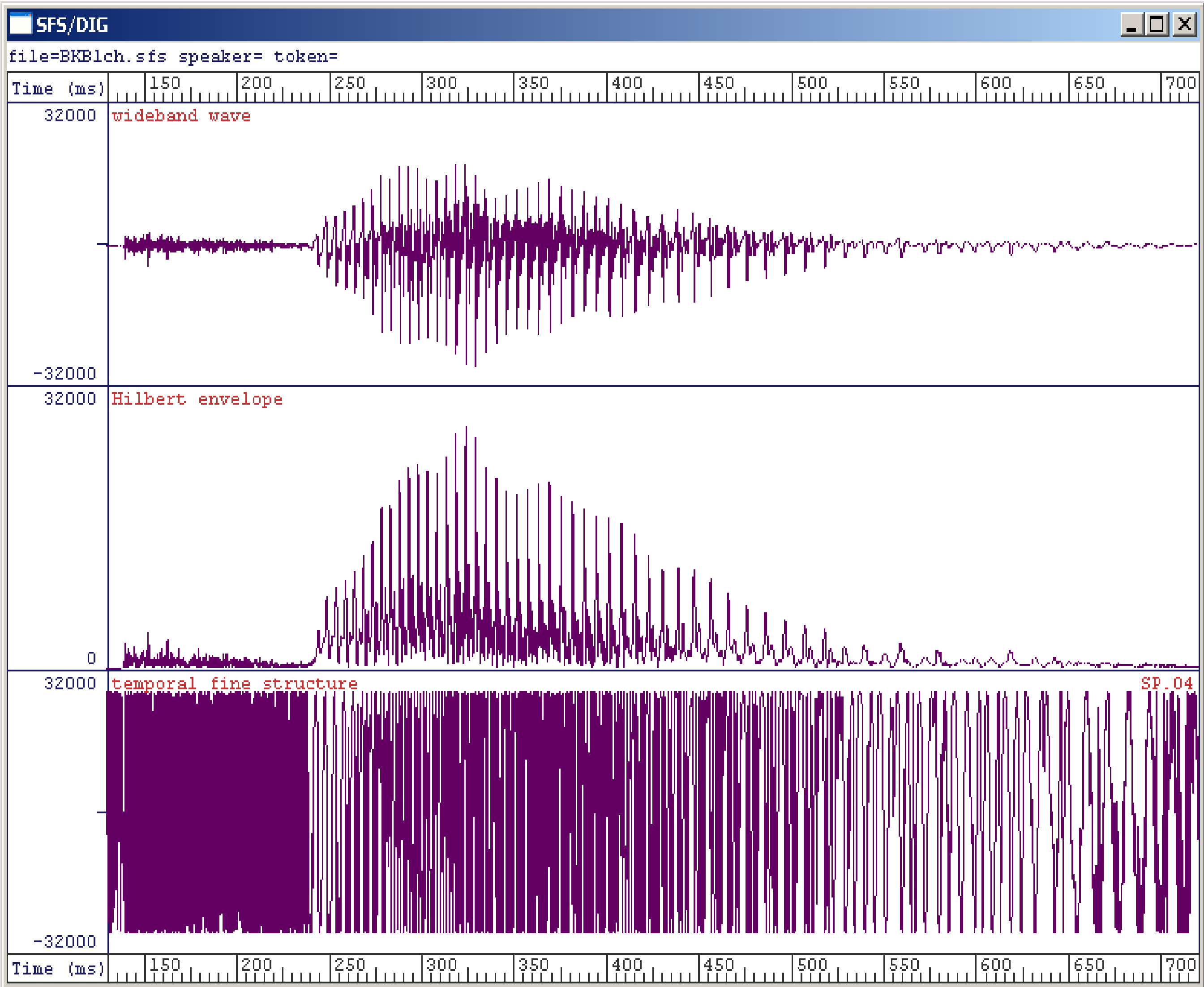
=

envelope

x

TFS

Decomposing a ‘clown’



original wave

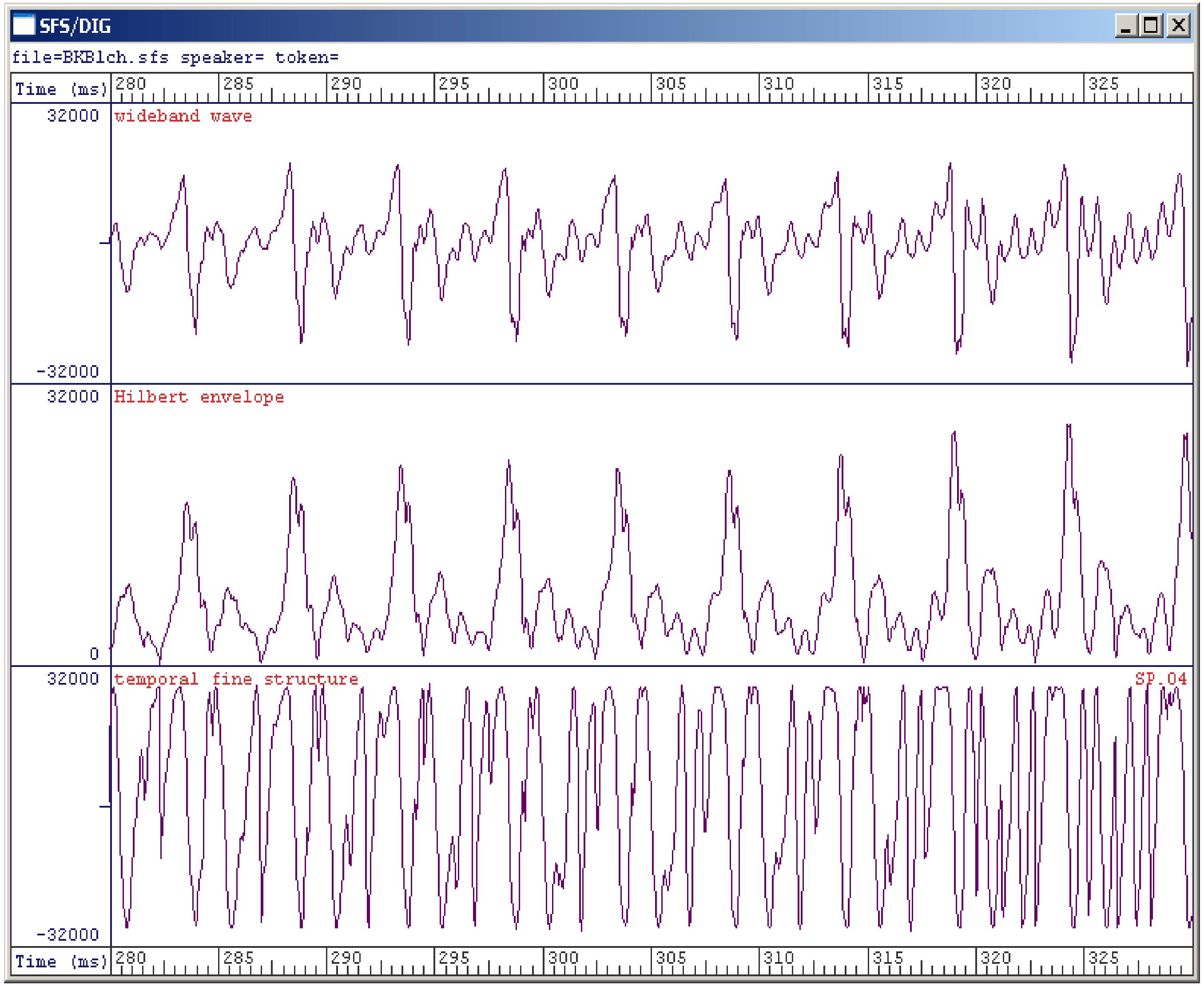
=

envelope

x

TFS

Look up close



original wave

=

envelope

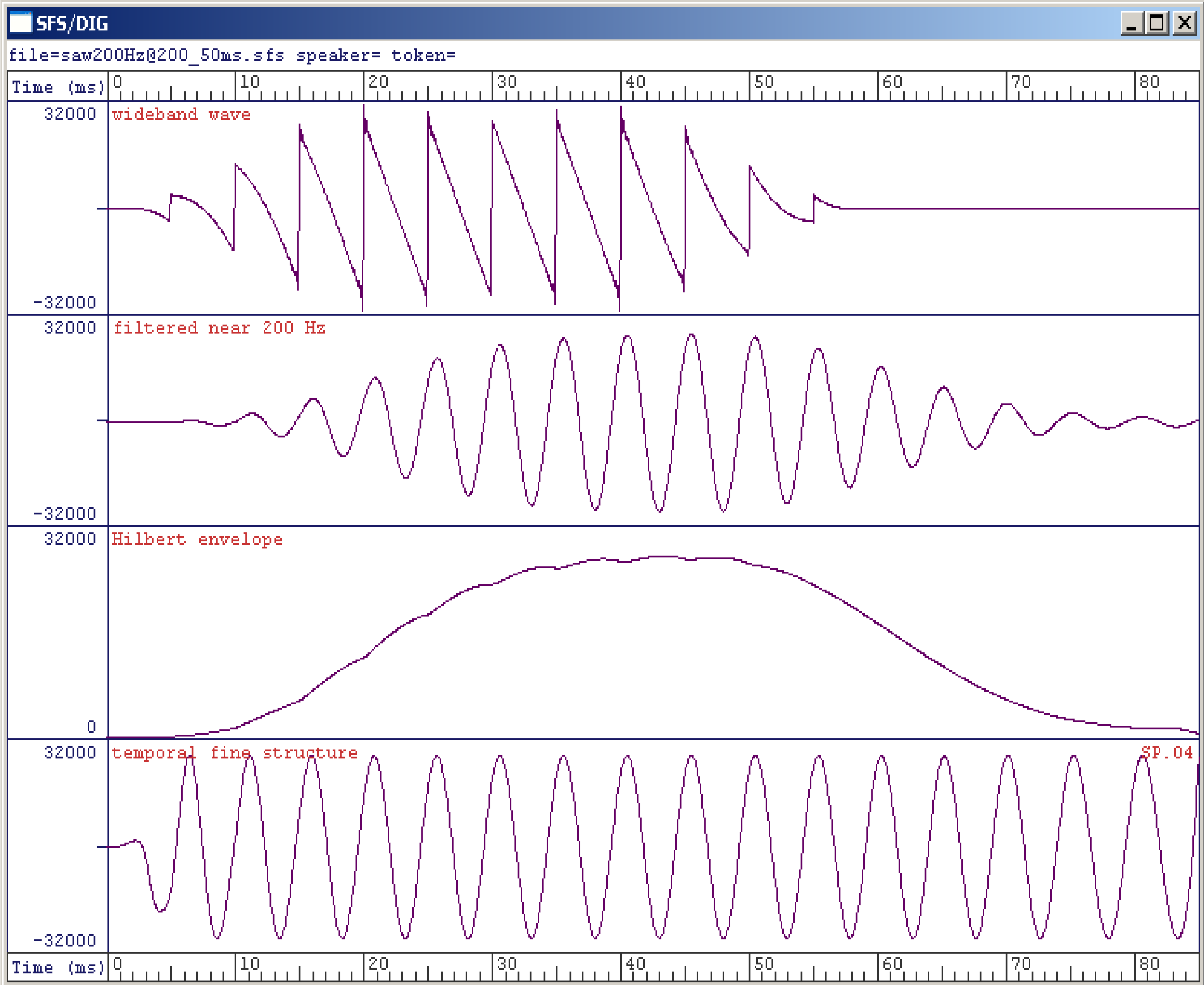
x

TFS
(hardly a ‘sinusoid’!)

A complication

- The auditory periphery acts as a kind of a filter bank
- So auditory nerve fibres transmit information about a bandpass filtered version of the original wide-band wave
- It only makes sense to apply the decomposition to a bandpass filtered version of the original wave
- Filter bandwidth will depend on
  - whether a listener is hearing-impaired
  - frequency in normal and hearing-impaired listeners
  - whether a listener is using a cochlear implant

Sawtooth: auditory filtering @ 200 Hz



original wave

filtered wave

=

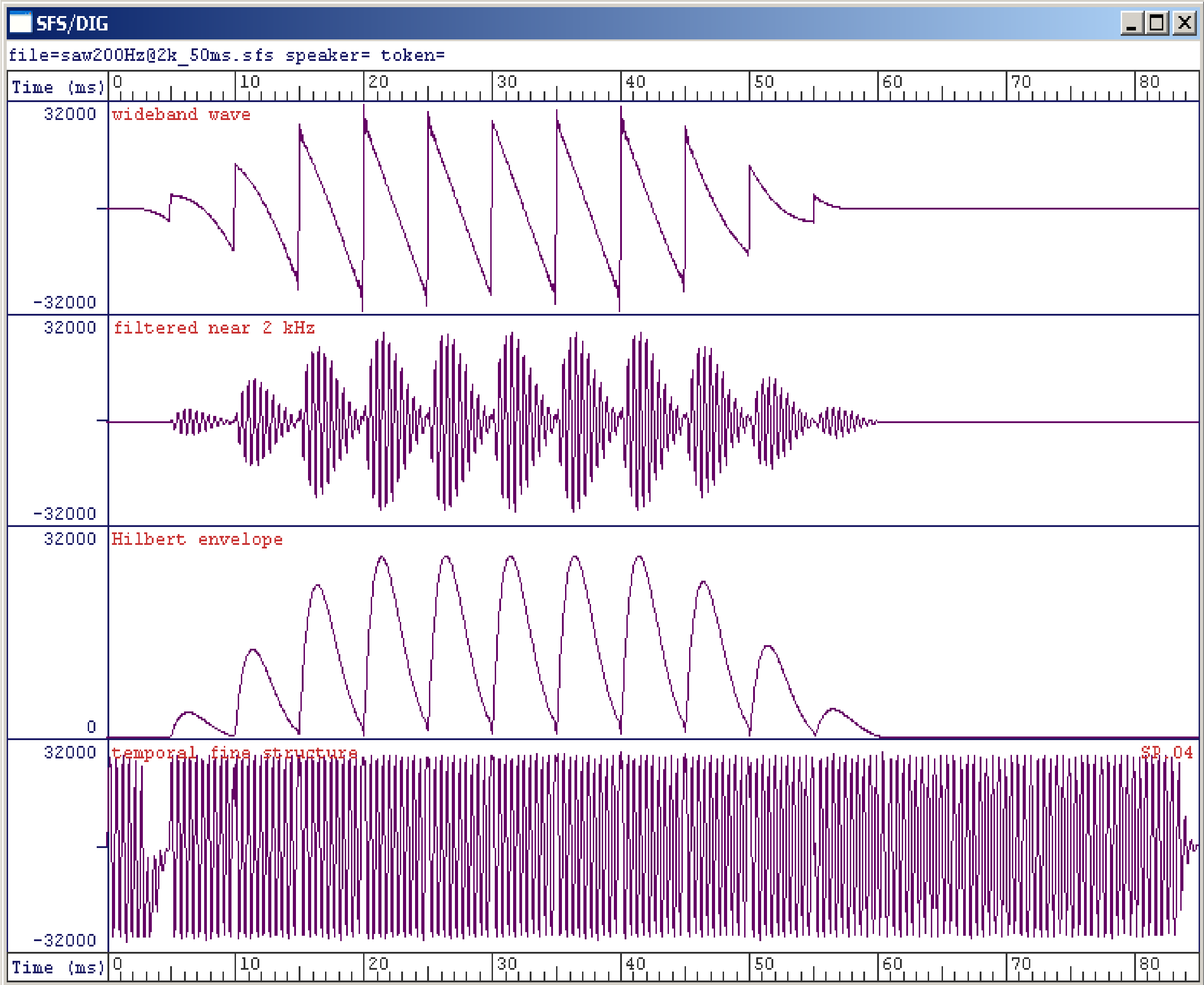
envelope

x

TFS

resolved harmonics — no evidence of periodicity in envelope; strong in TFS

Sawtooth: auditory filtering @ 2 kHz



original wave

filtered wave

=

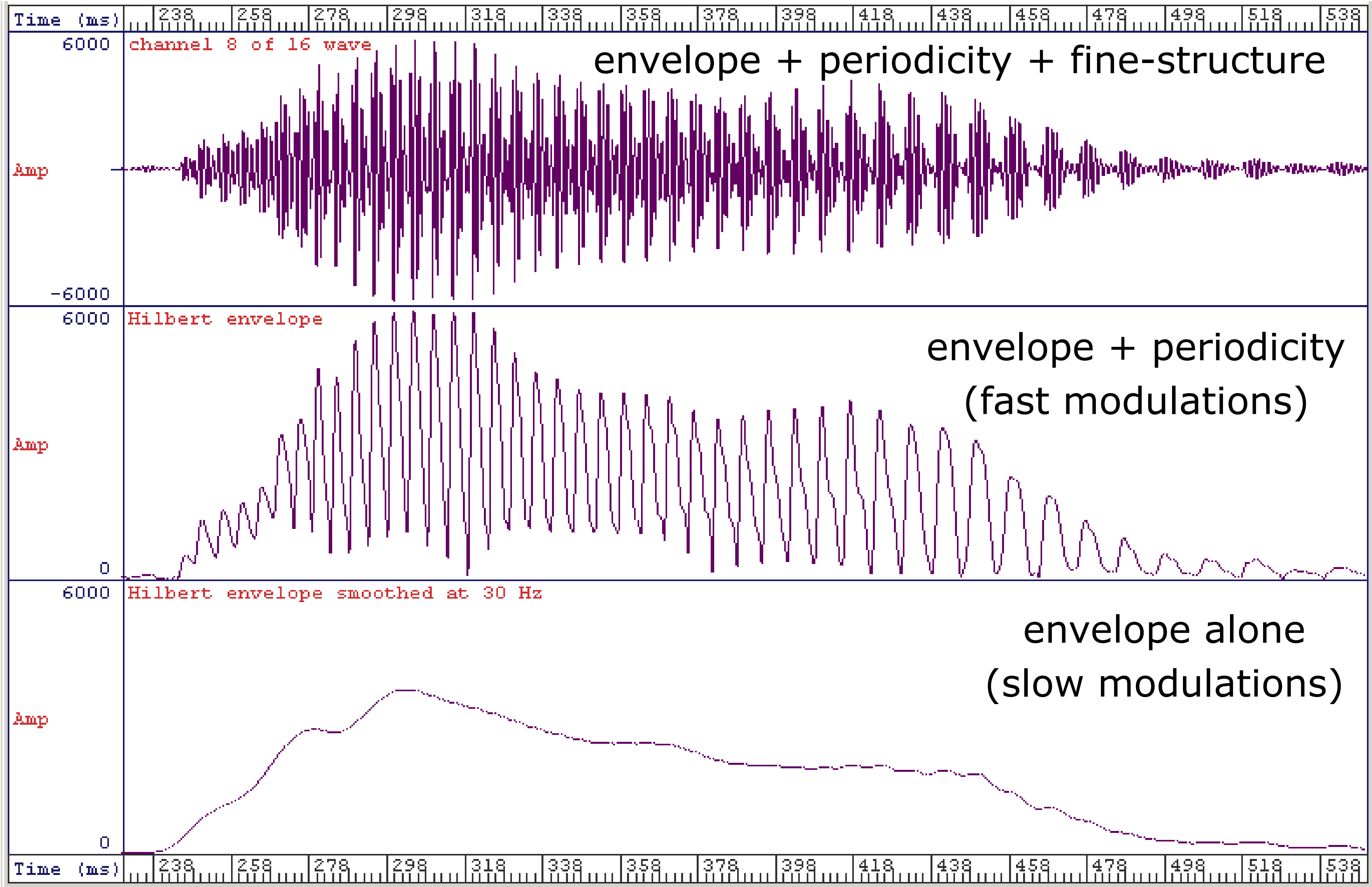
envelope

x

TFS

unresolved harmonics — periodicity evident in envelope; weak in TFS

A 3-way partition of temporal information



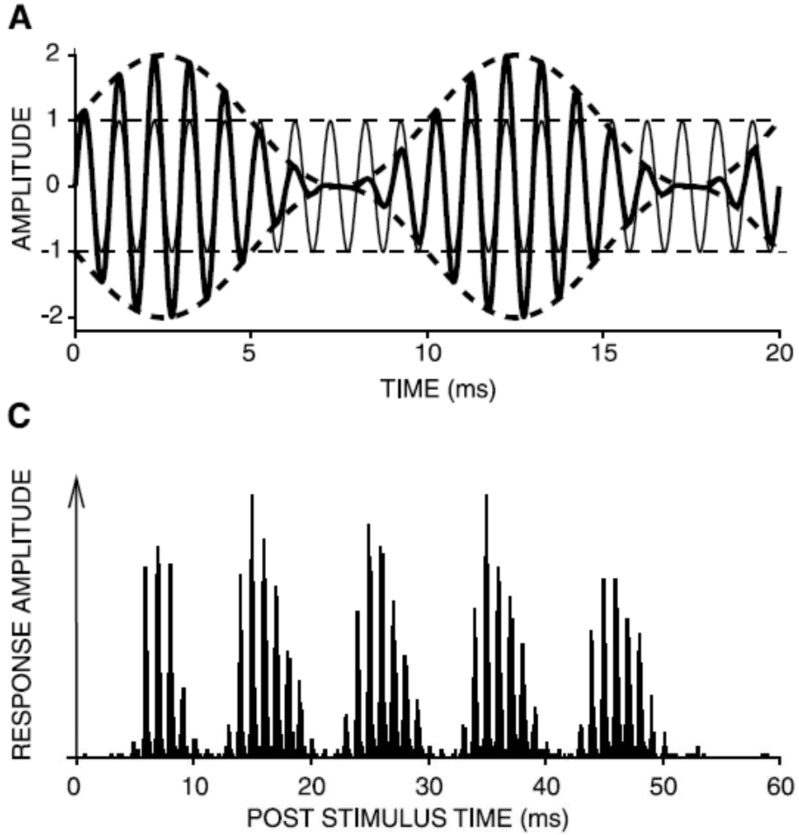


FIG. 1. *A*: superimposed waveforms of an unmodulated 1,000-Hz tone (thin line) and the same tone sinusoidally amplitude modulated (AM) (thick line) at 100% with a modulation frequency of 100 Hz, according to *Equation 1*. Dashed lines indicate the envelope. The amplitude is referenced to the peak amplitude of the unmodulated tone. *B*: idealized spectrum of the AM tone in *A*. At 100% modulation, the amplitude of the sidebands is half that of the carrier, i.e., a difference of 6 dB. *C*: average response in the form of a poststimulus time (PST) histogram of a nerve fiber to the signal shown in *A* (stimulus duration, 50 ms). *D*: spectrum of the PST histogram in *C*. The components at carrier frequency (f_c) and $f_c \pm$ modulation frequency (f_m) indicate that there is phase-locking to the fine-structure of the stimulus waveform. The component at f_m is prominently present in the response but is absent in the stimulus (*B*). The small circle on the ordinate indicates the average firing rate.

All 3 temporal features preserved in the auditory nerve

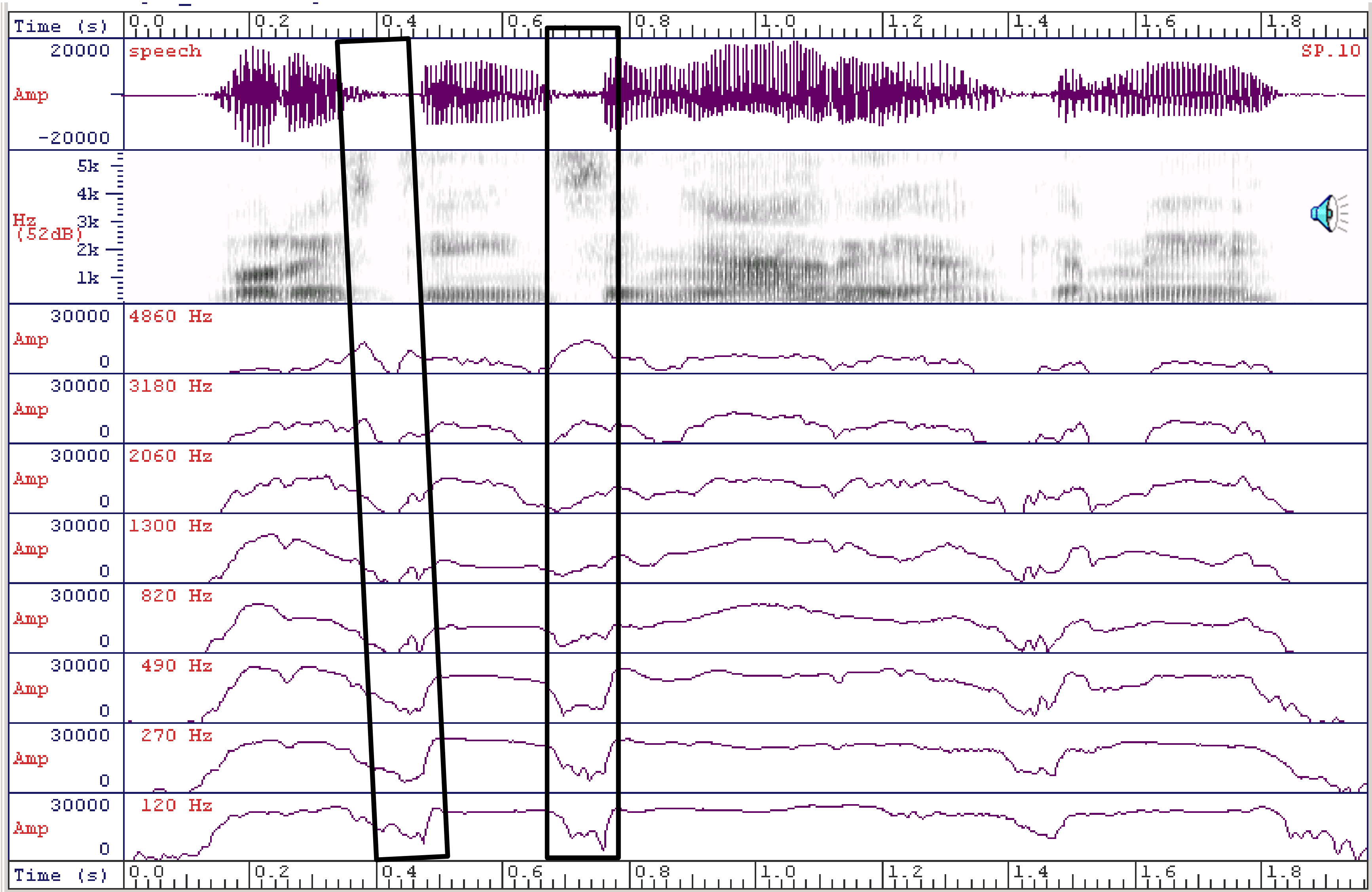
(slower modulations not shown)

Joris *et al.* 2004

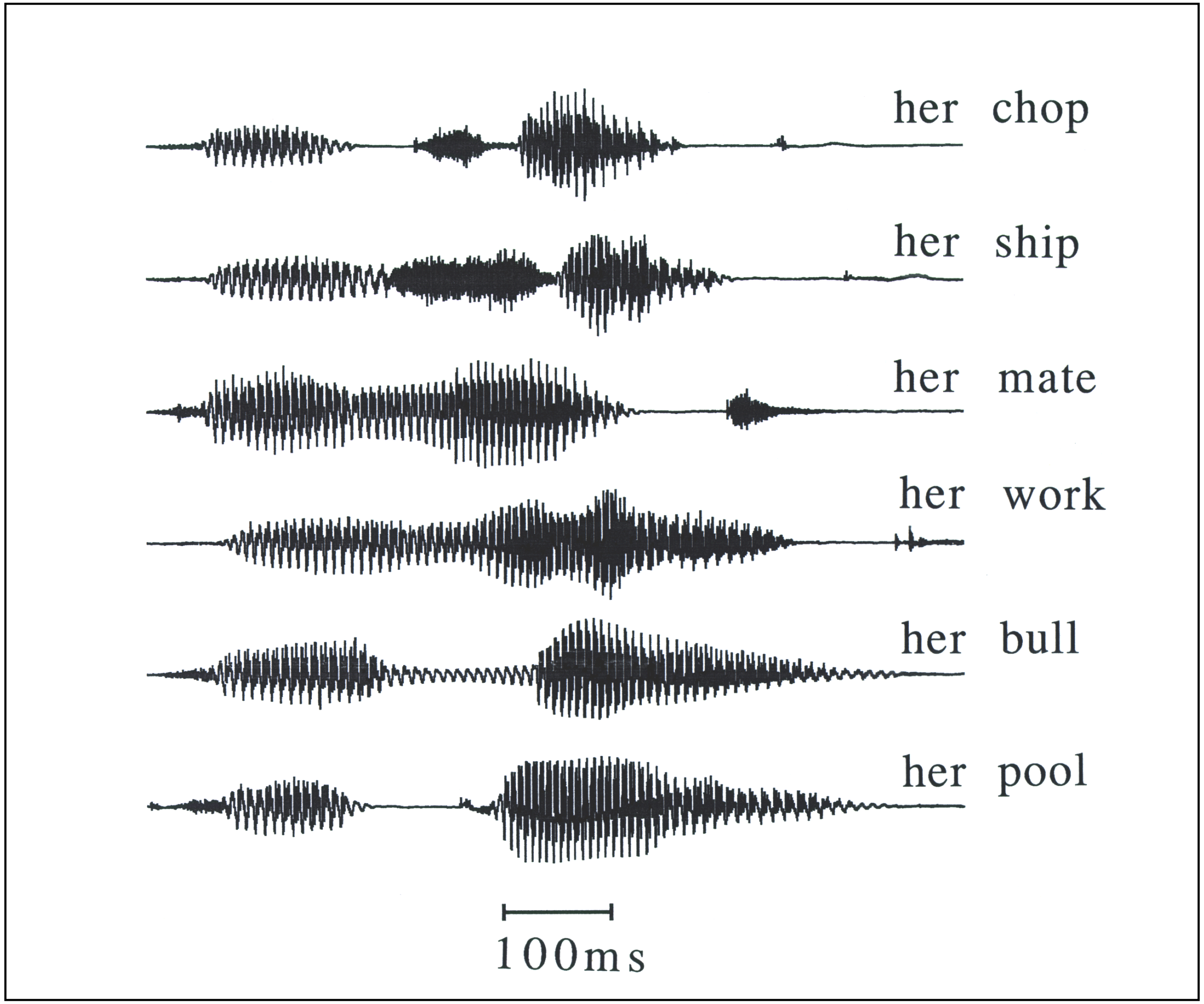
Everyone agrees that …

- ‘Slowish’ envelopes (<30 Hz or so) are really important for speech perception
- Distinguish two features
  - Envelope variations that are highly correlated across frequency
  - And those that are not.

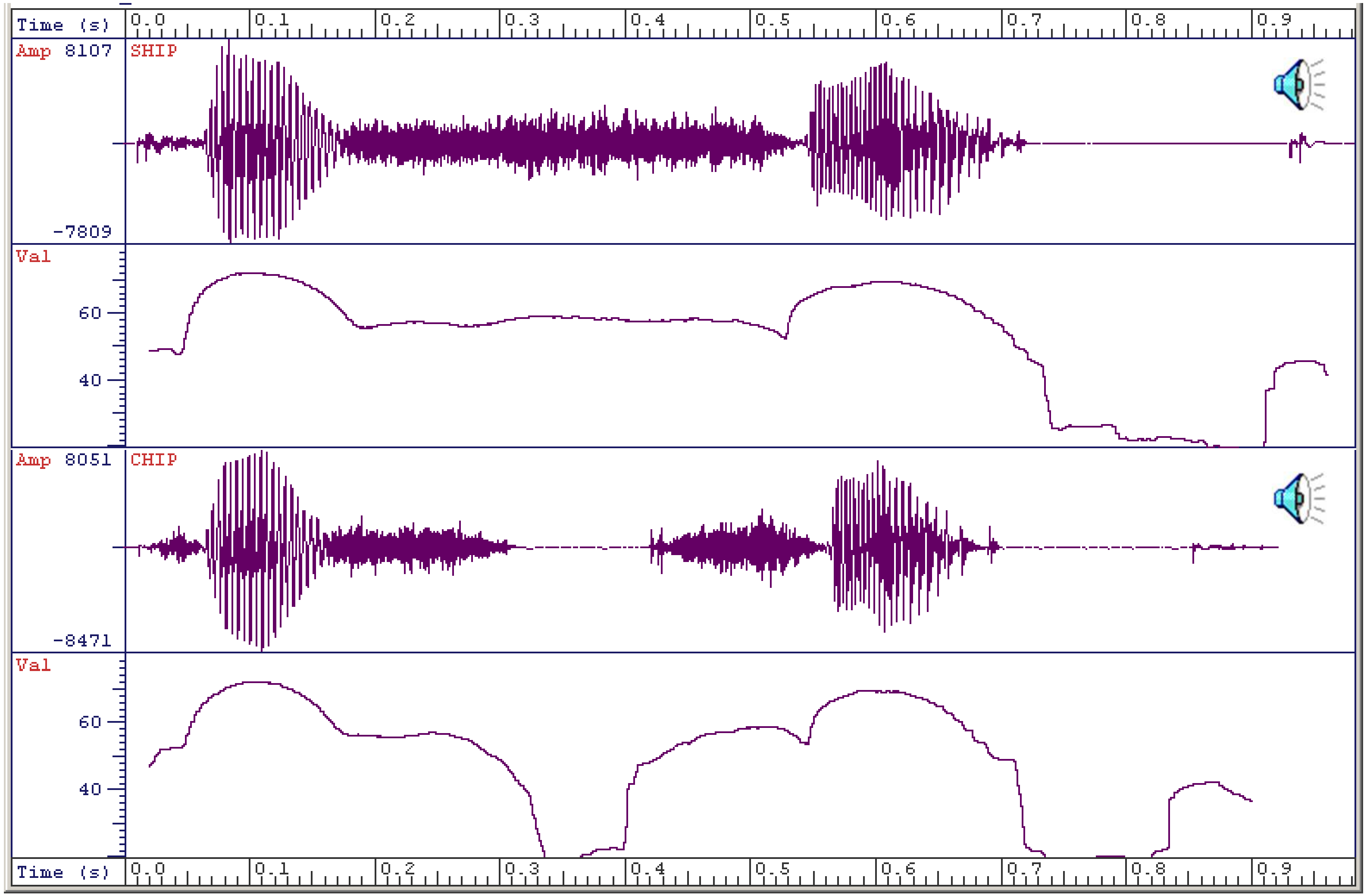
Correlated and uncorrelated (across frequency) envelope modulations

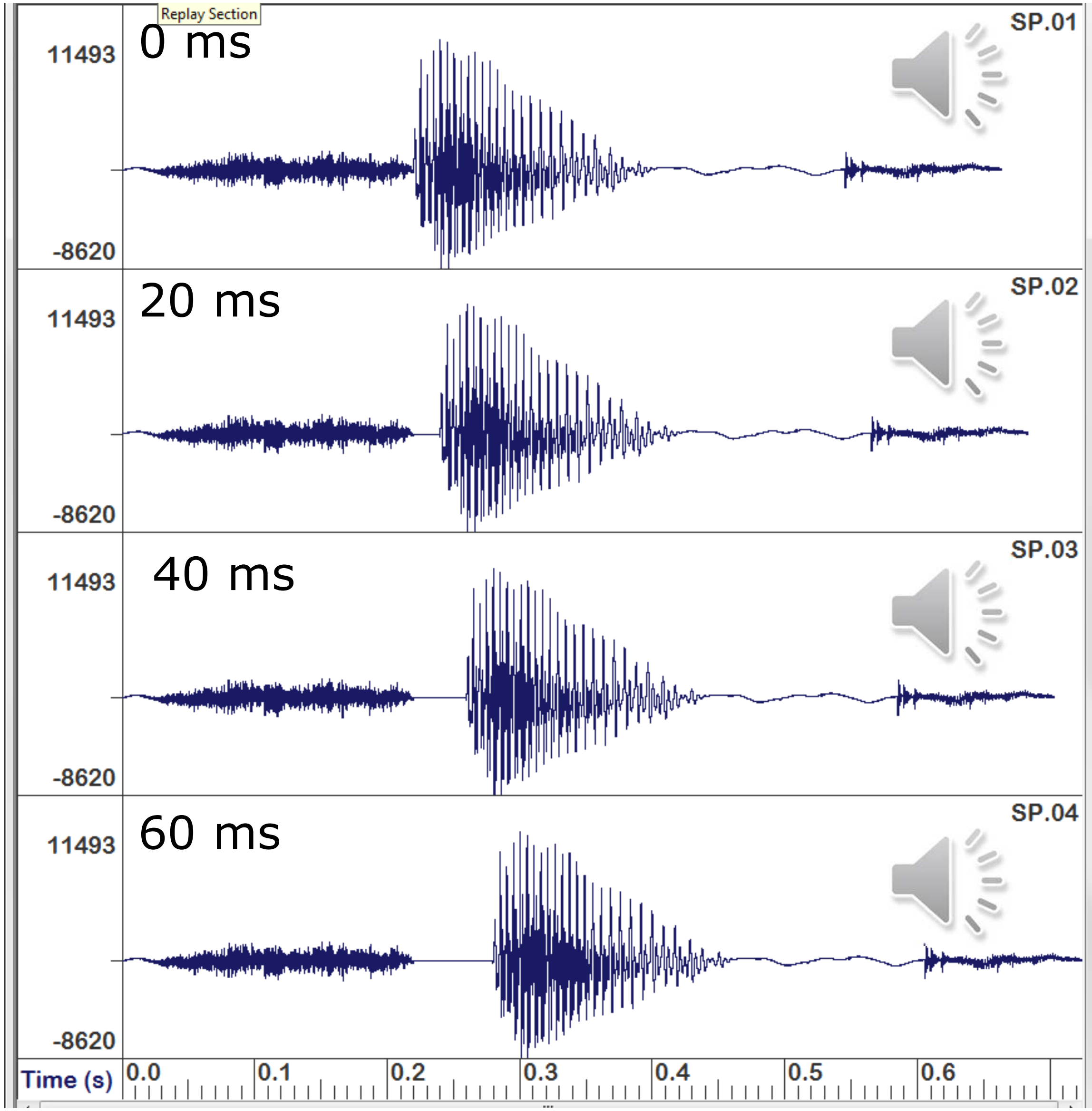


Correlated envelopes in speech – one source of cues to consonants

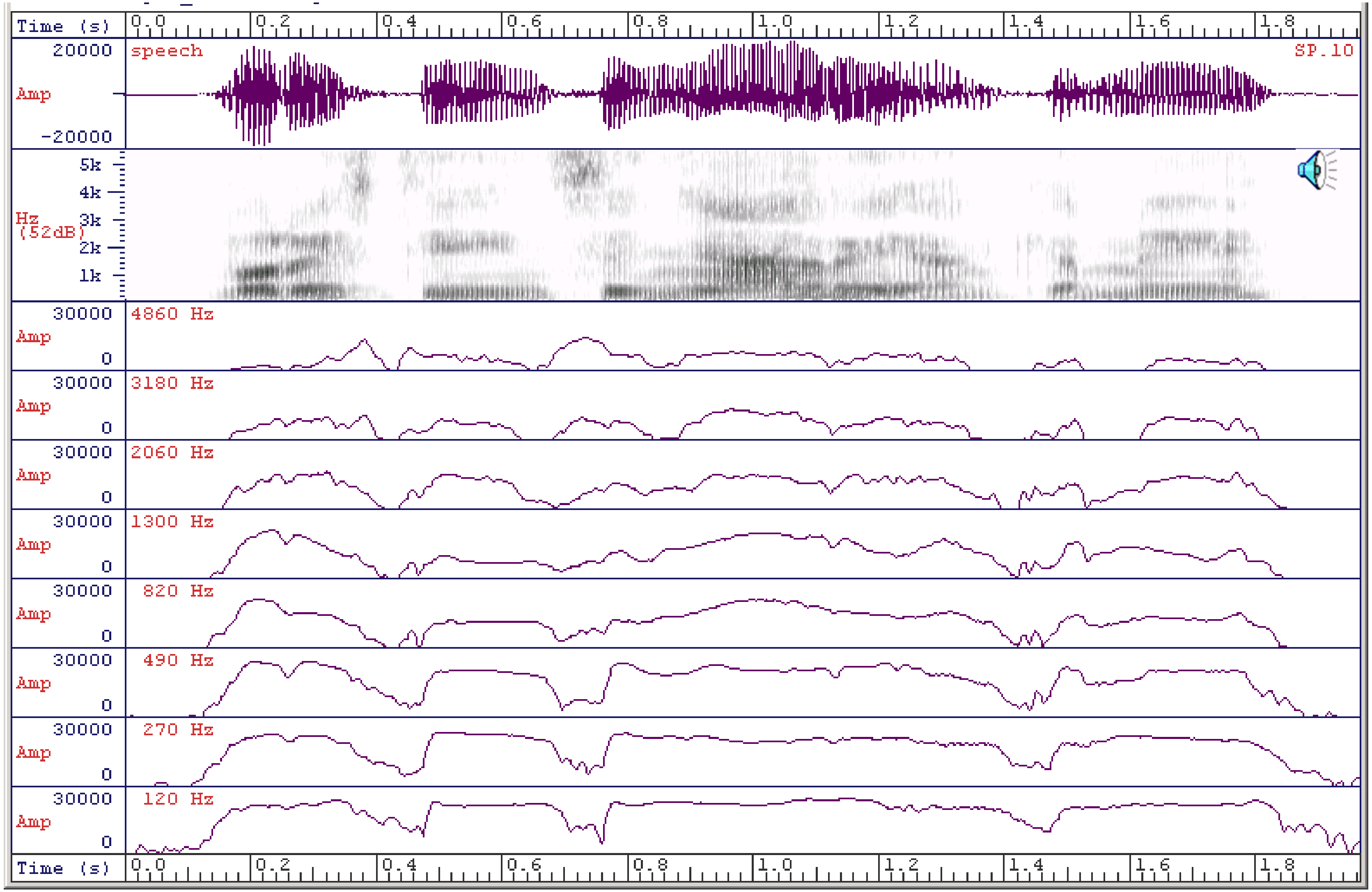


Changing manner of articulation
push ship vs. *push chip*

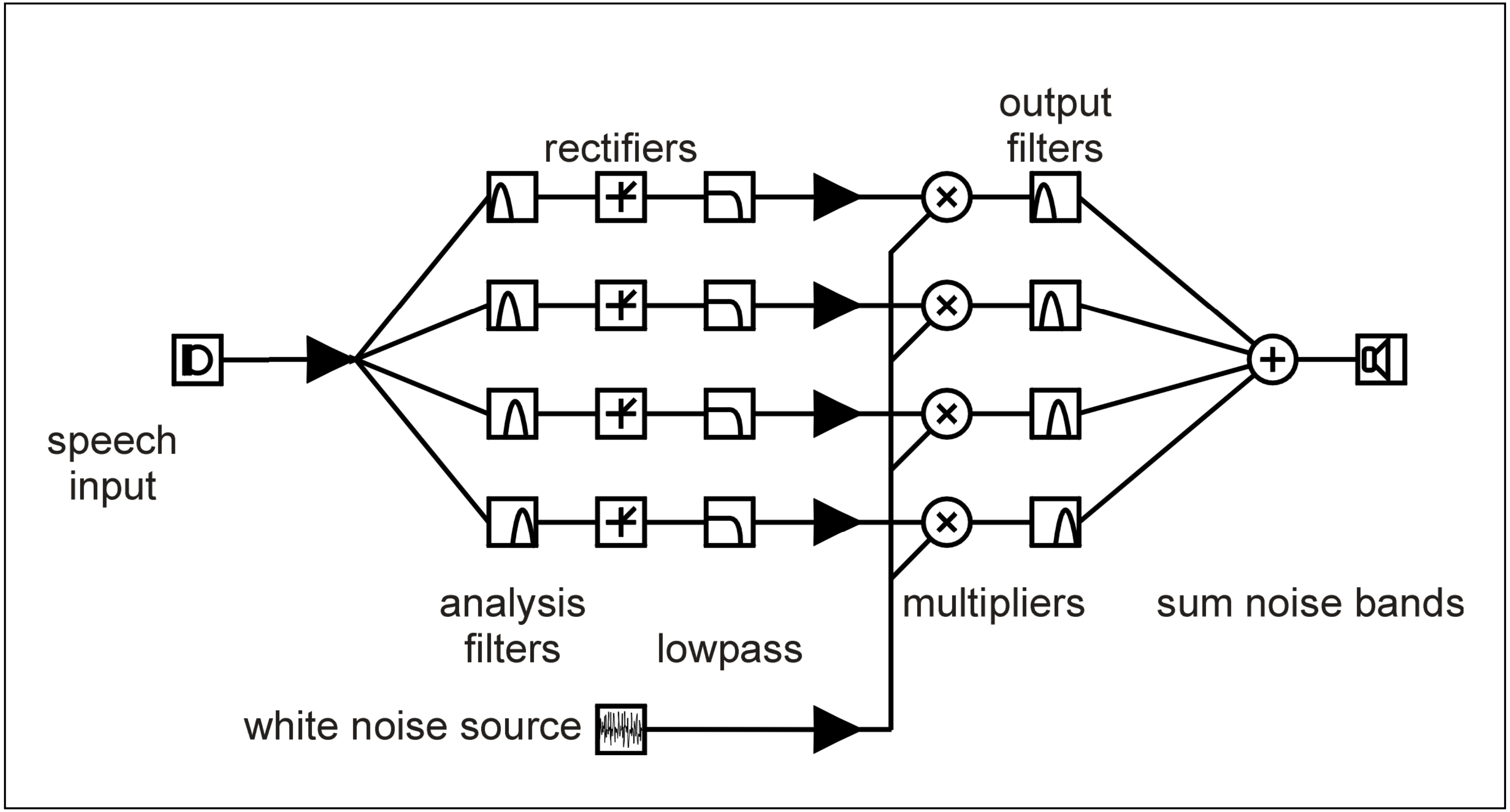



Replay Section
0 ms
20 ms
40 ms
60 ms
SP.01
SP.02
SP.03
SP.04
11493
-8620
Time (s)
0.0
0.1
0.2
0.3
0.4
0.5
0.6

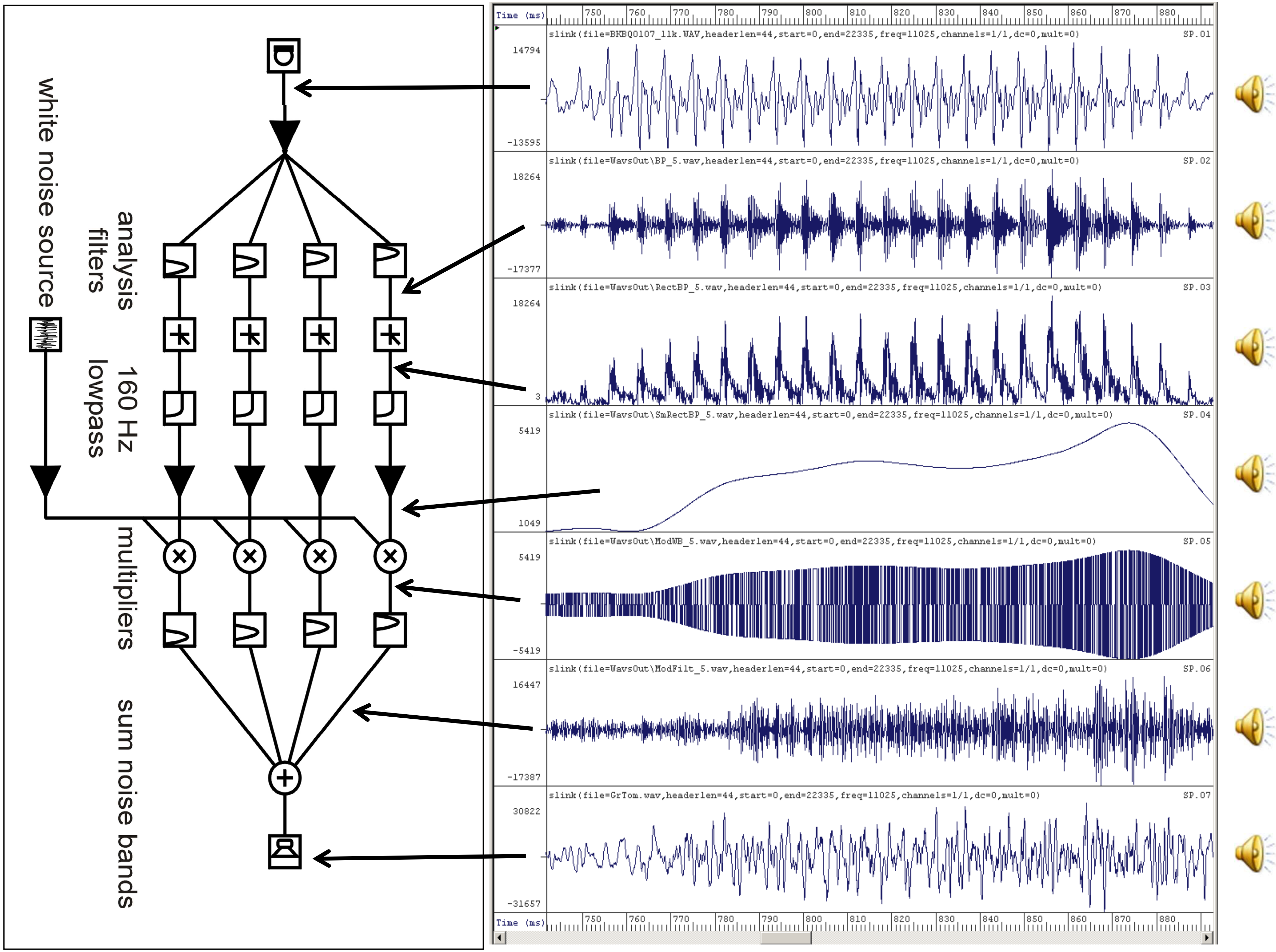
Spectral dynamics are encoded in uncorrelated across-channel envelope modulations



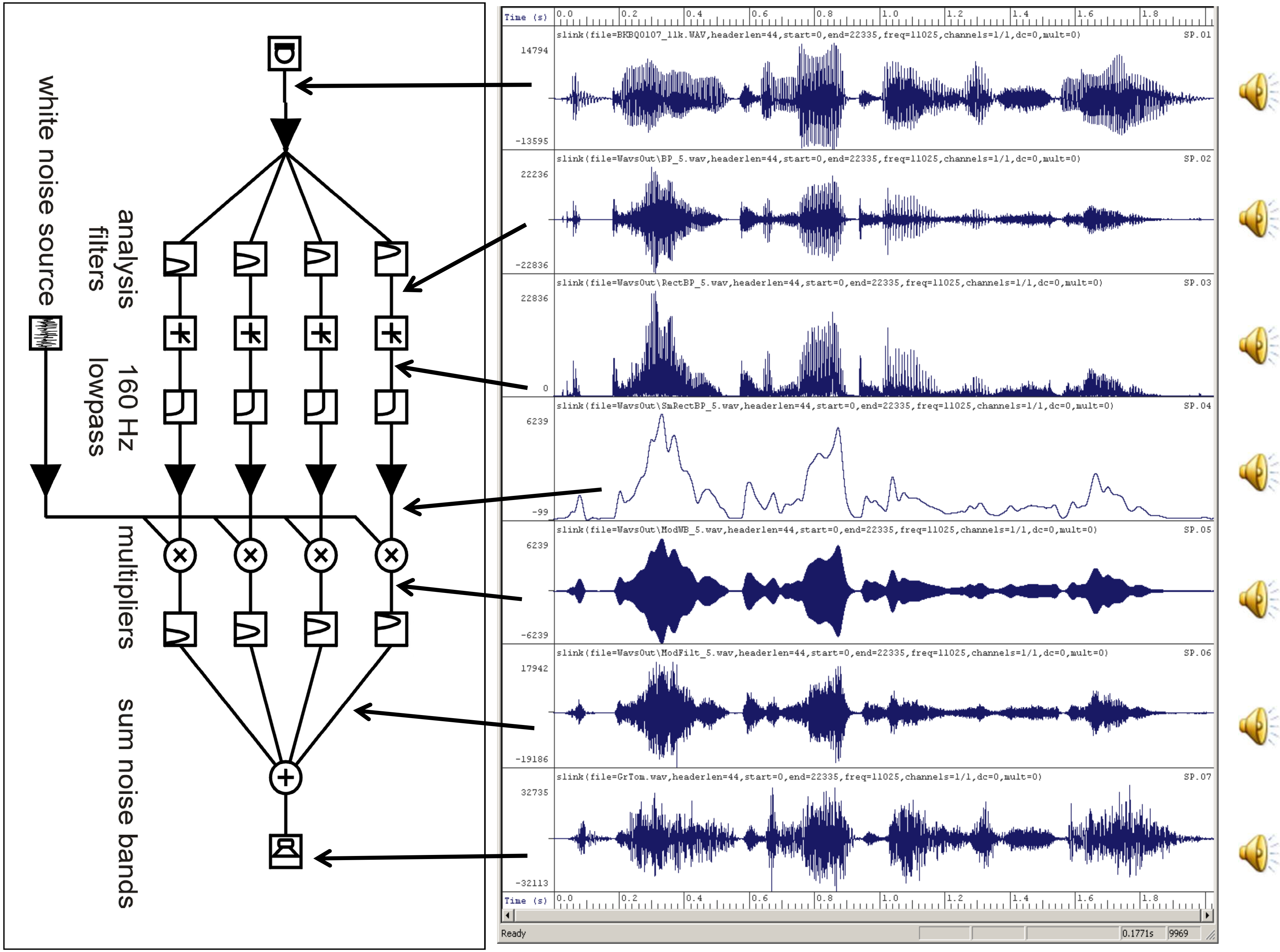
Proof that envelopes are sufficient: Noise-excited vocoding

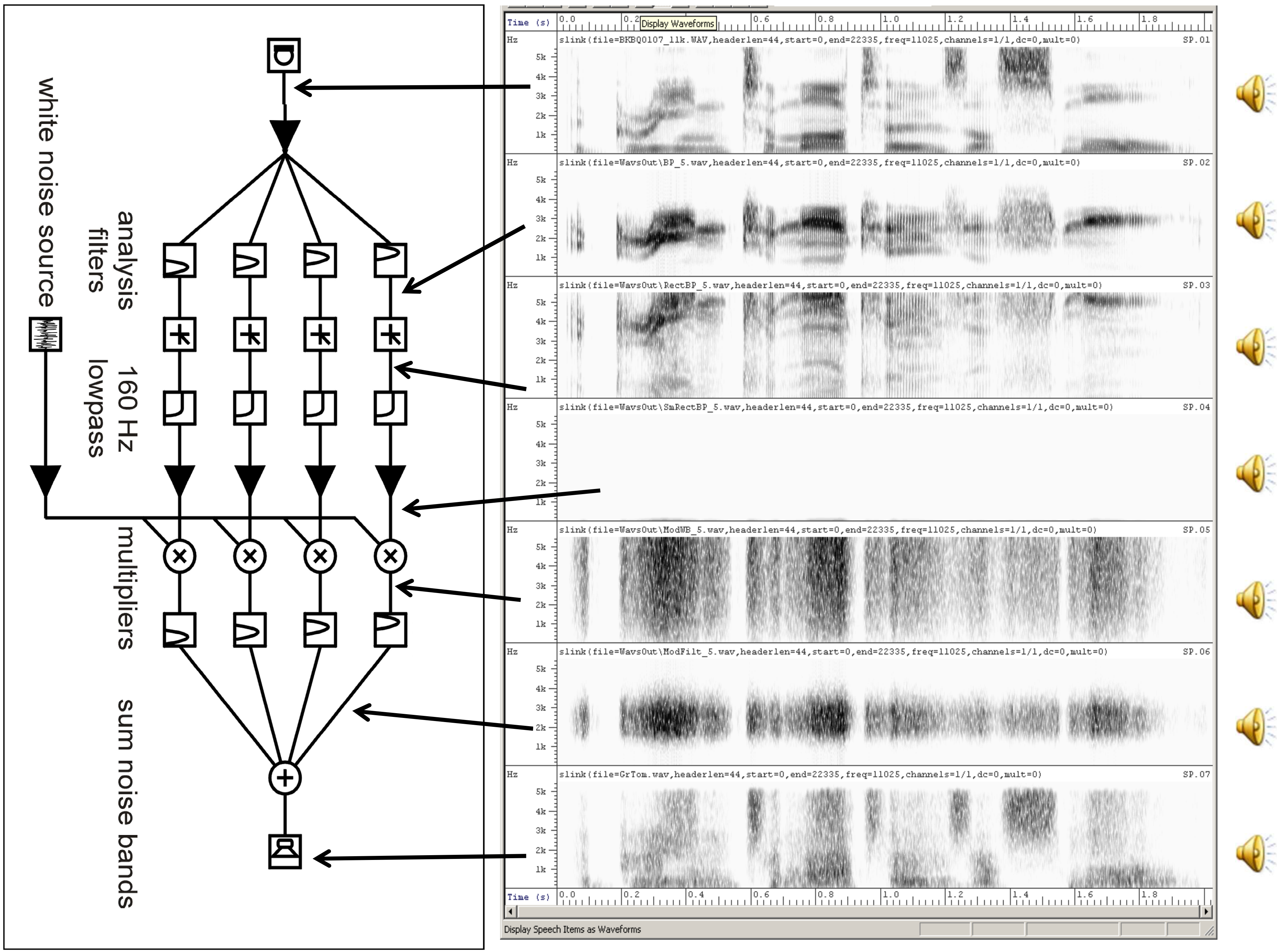


more or less preserves envelopes, destroys TFS



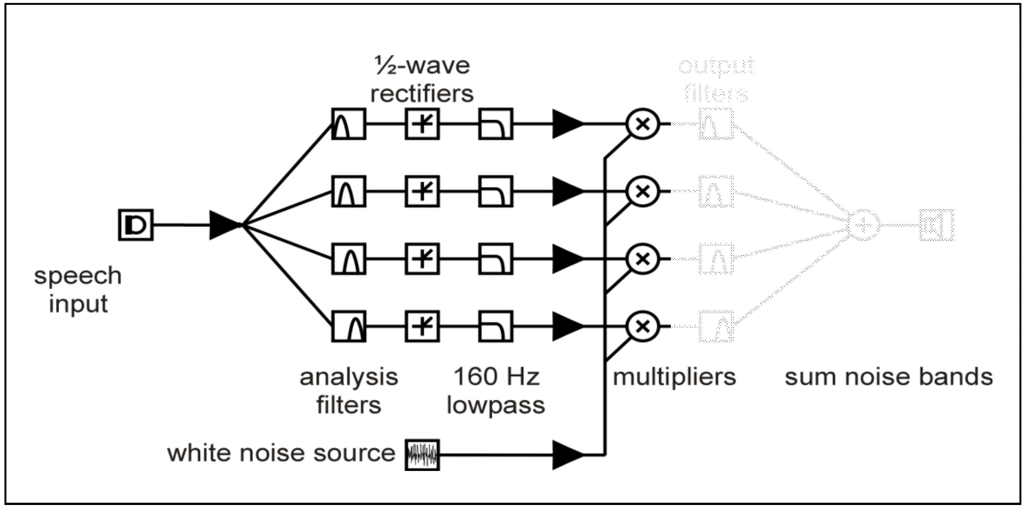
white noise source
analysis filters
160 Hz lowpass
multipliers
sum noise bands
Time (ms)
slink(file=BKBQ0107_11k.WAV,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.01
slink(file=WavsOut\BP_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.02
slink(file=WavsOut\RectBP_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.03
slink(file=WavsOut\SmRectBP_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.04
slink(file=WavsOut\ModWB_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.05
slink(file=WavsOut\ModFilt_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.06
slink(file=GrTom.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.07

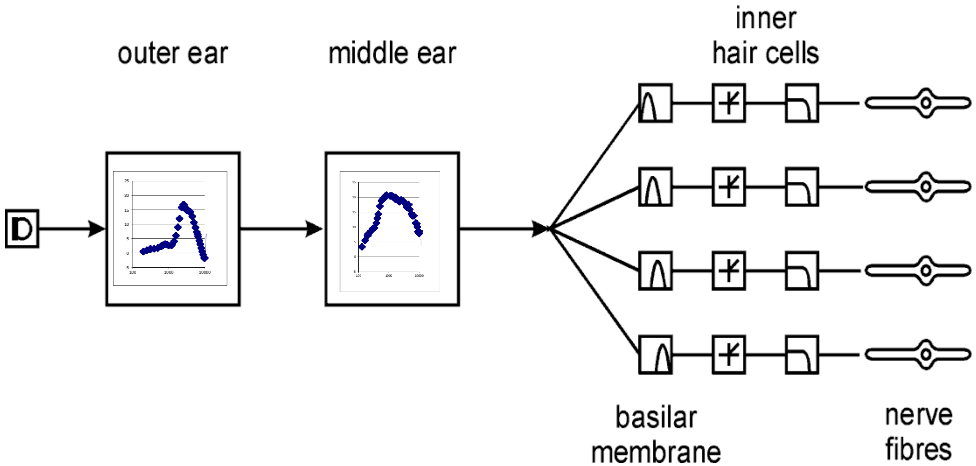

white noise source
analysis filters
160 Hz lowpass
multipliers
sum noise bands
Time (s)
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8
slink(file=BKBQ0107_11k.WAV,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.01
14794
-13595
slink(file=WavsOut\BP_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.02
22236
-22836
slink(file=WavsOut\RectBP_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.03
22836
0
slink(file=WavsOut\SmRectBP_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.04
6239
-99
slink(file=WavsOut\ModWB_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.05
6239
-6239
slink(file=WavsOut\ModFilt_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.06
17942
-19186
slink(file=GrTom.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.07
32735
-32113
Time (s)
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8
Ready
0.1771s
9969



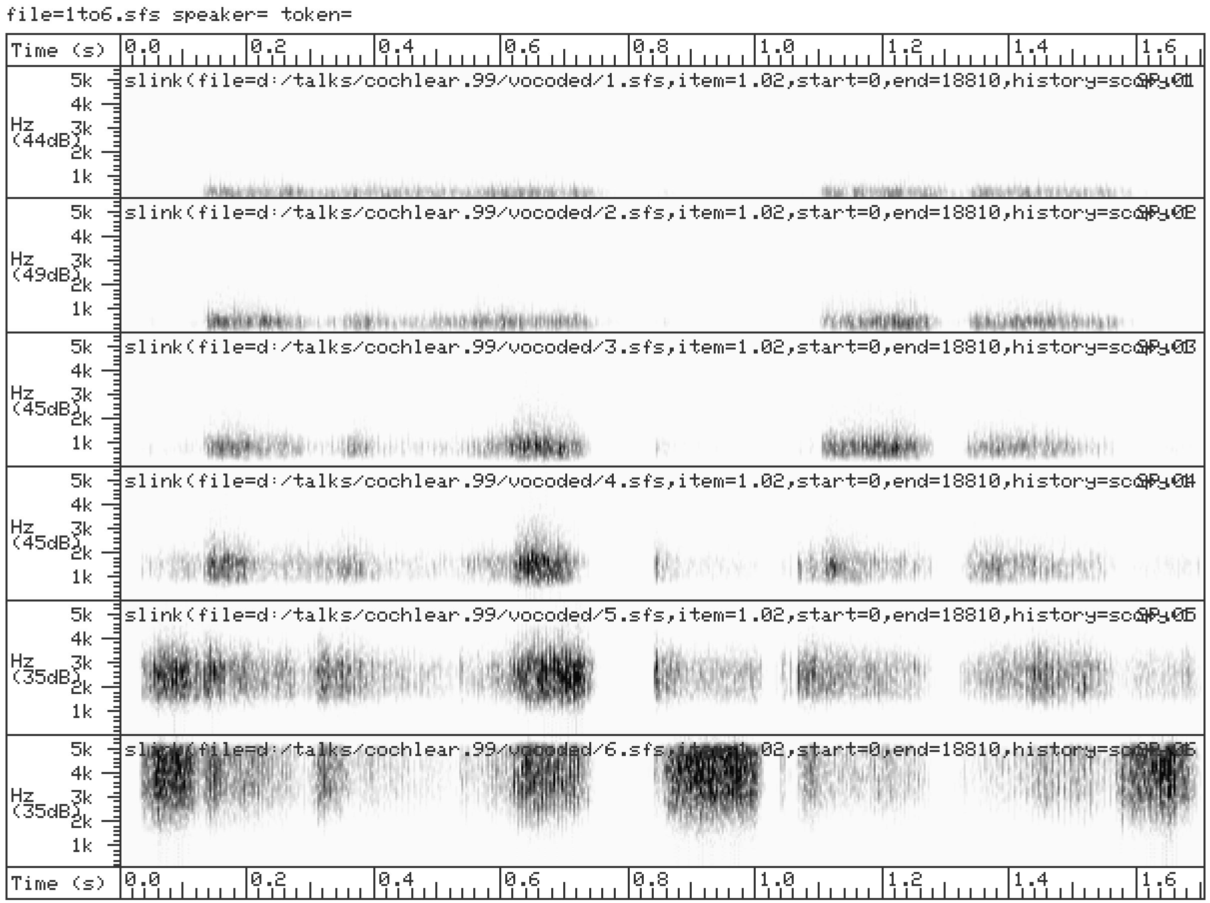
white noise source
analysis filters
160 Hz lowpass
multipliers
sum noise bands
Time (s)
Display Waveforms
slink(file=BKBQ0107_11k.WAV,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.01
slink(file=WavsOut\BP_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.02
slink(file=WavsOut\RectBP_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.03
slink(file=WavsOut\SmRectBP_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.04
slink(file=WavsOut\ModWB_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.05
slink(file=WavsOut\ModFilt_5.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.06
slink(file=GrTom.wav,headerlen=44,start=0,end=22335,freq=11025,channels=1/1,dc=0,mult=0)
SP.07
Display Speech Items as Waveforms

Note similarity to normal cochlear processing

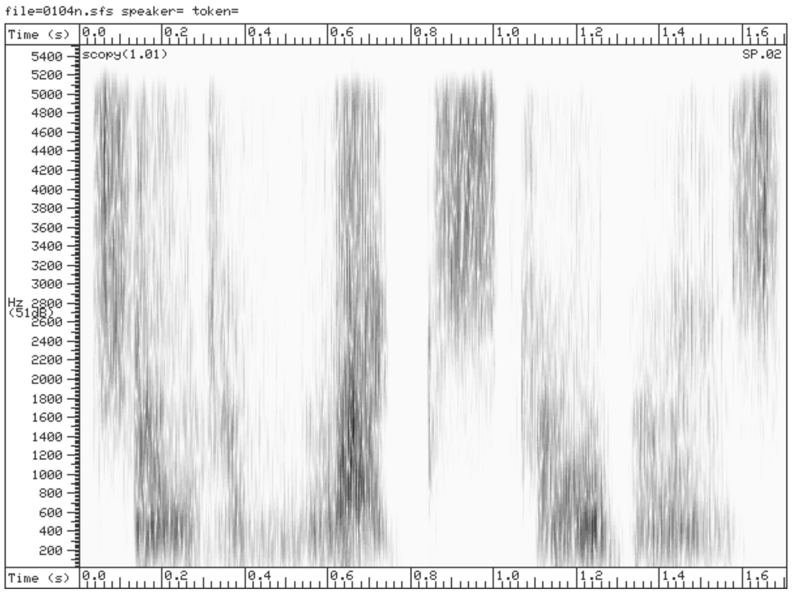


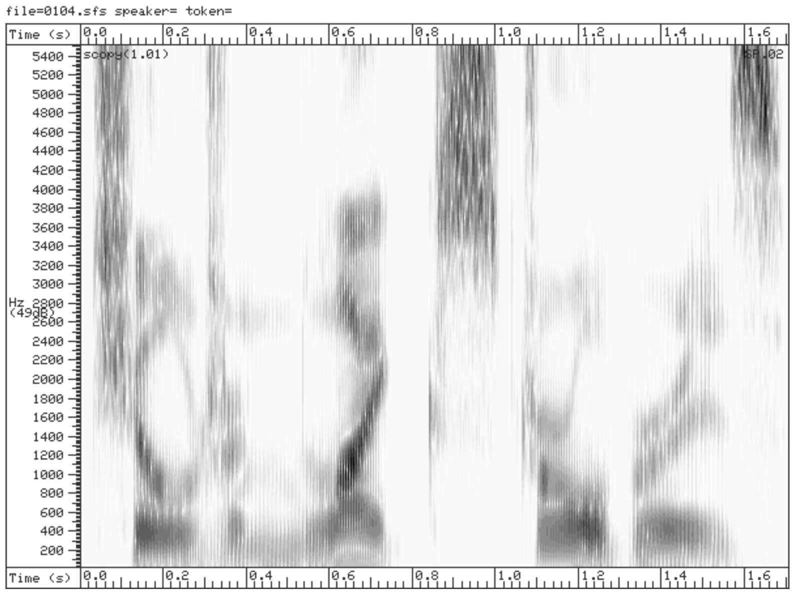


Separate channels in a 6-channel simulation

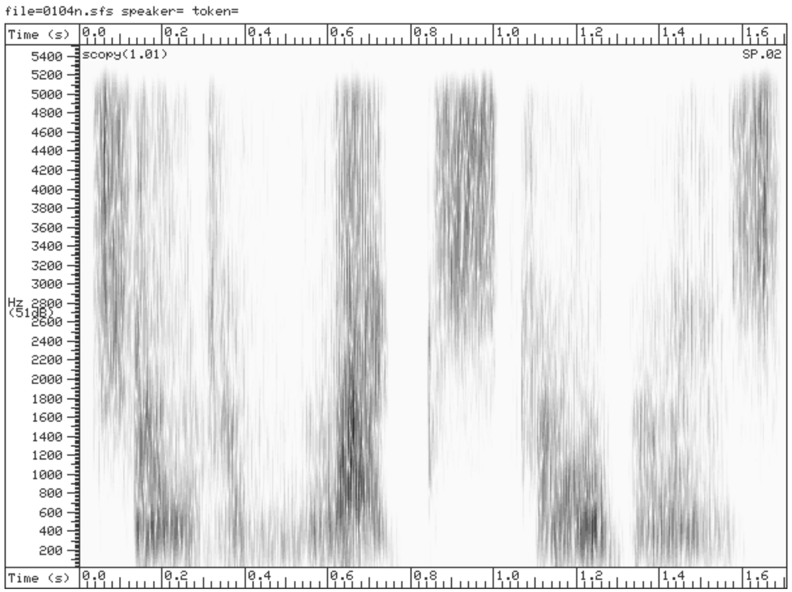


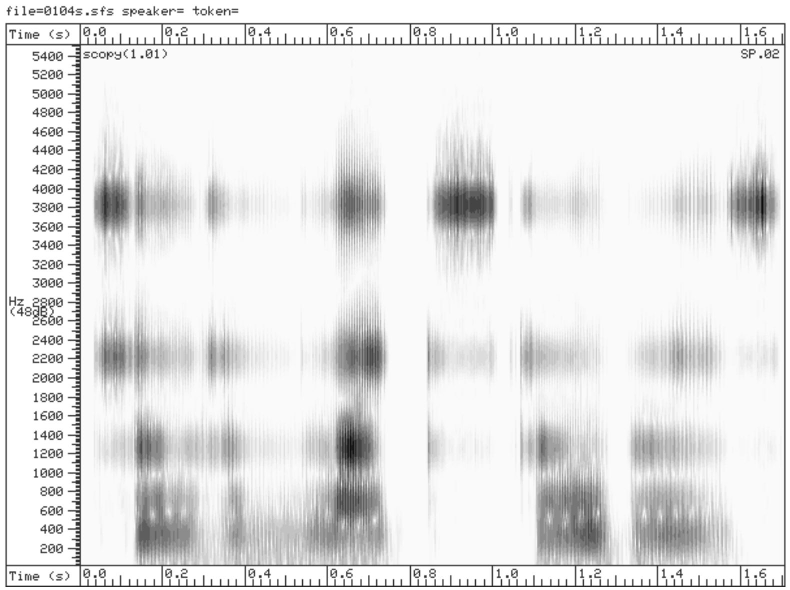
... and when summed together.



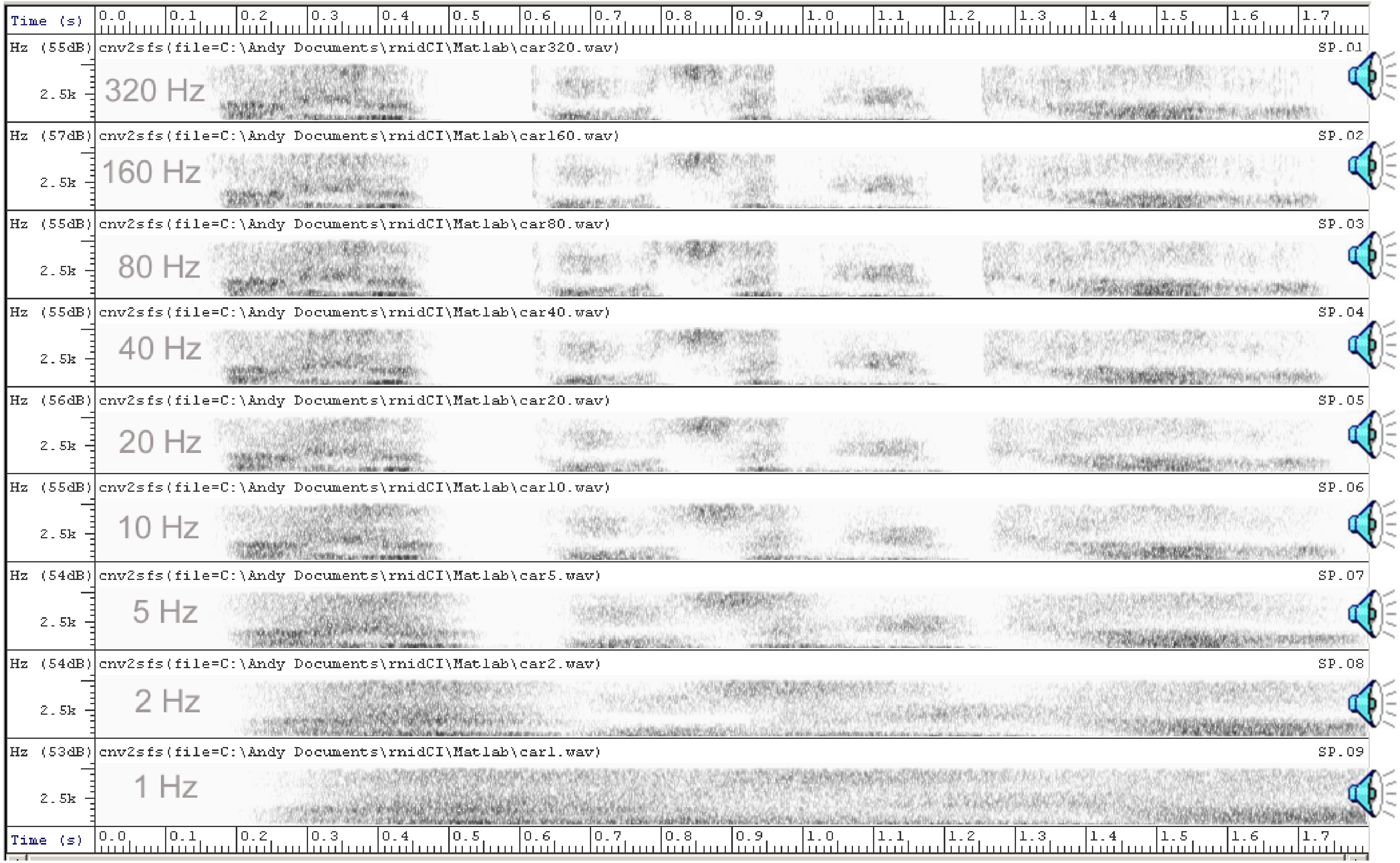


Never mind the quality...
feel the intelligibility.





Effects of envelope smoothing on speech - modulations below 10 Hz are most important



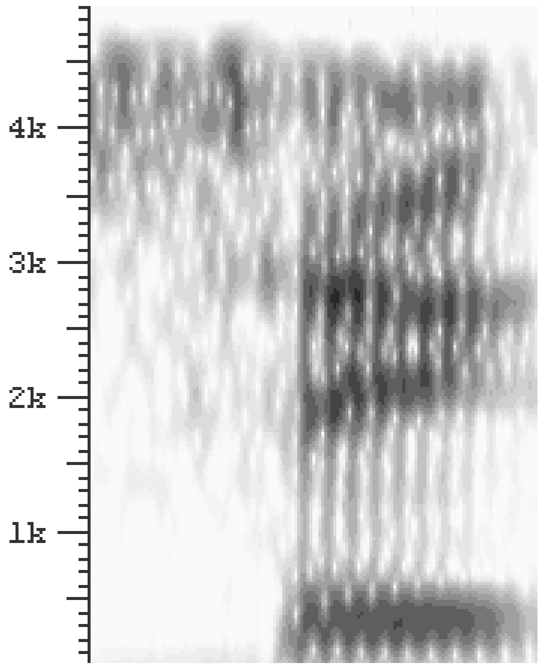
So what’s missing in envelope?

- TFS *is* important for …
  - Localisation
  - Perception of melodic pitch
    - Intonation and tone, for the TFS of a periodic sound
- In CI research, TFS often used as a code word for ‘pitch perception’
  - Even though poor pitch perception may also arise from impaired frequency selectivity.

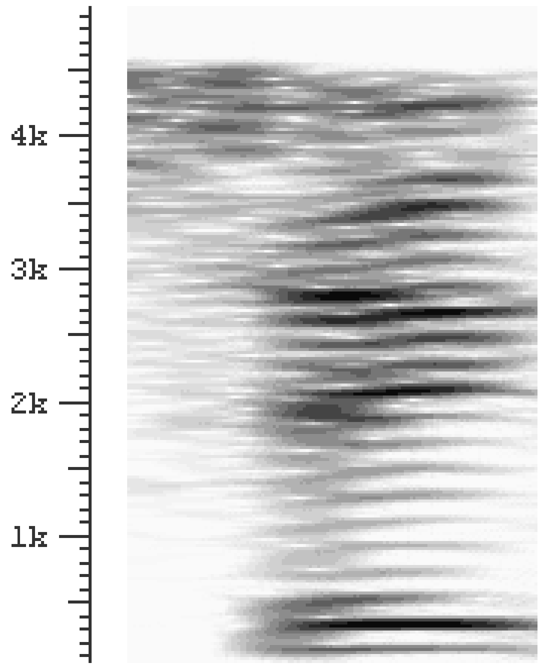
NHLs do use TFS for pitch

Types of Spectrogram

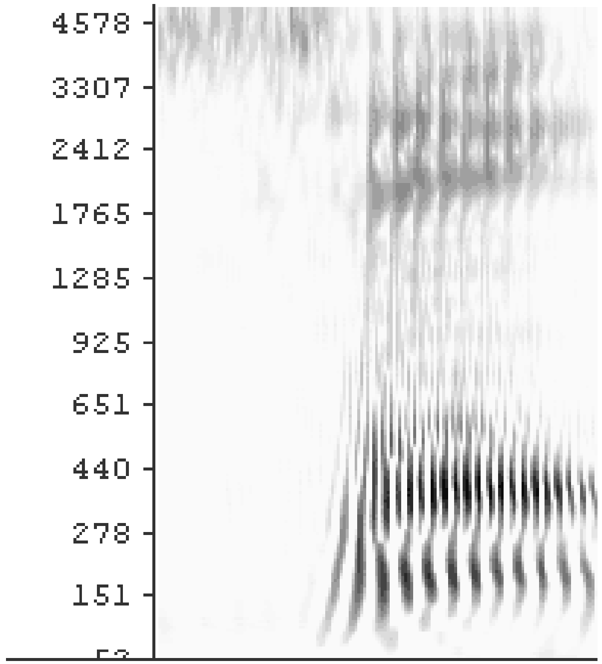
Wide-band



Narrow-band



Auditory



An auditory spectrogram looks like a wide-band spectrogram at high frequencies and a narrow-band spectrogram at low frequencies (but with more temporal structure).

Summary

- Waveforms (after any filter bank/spectral analysis) can be decomposed into the product of
  - An envelope (something fairly slow)
    - o often divisible into slower and faster components
  - A TFS (something fast)
- Envelope is necessary and sufficient for speech perception in quiet
- One serious limitation of CIs (and HI listeners) especially for speech in noise may be poor access to TFS information
  - But the representation of TFS also depends upon frequency selectivity, so it is not necessarily easy to separate out their effects