AUDL 4007 & GS12 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)



Adding waves (frequency domain)



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

based on *multiplication*.

Multiplying (modulating) waves



Multiplying (modulating) waves



Can work this backwards too



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

Extracting envelopes



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

 $x(t) = ENV(t) \cdot \sin[2\pi ft + \Theta(t)]$

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



A simple example: a tone pulse



A simple example: a noise pulse



A simple example: a sawtooth



Decomposing a 'clown'



Look up close



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



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

Sawtooth: auditory filtering @ 2 kHz



unresolved harmonics - periodicity evident in envelope; weak in TFS

A 3-way partition of temporal information





All 3 temporal features preserved in the auditory nerve (slower modulations not shown)

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.

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





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







Note similarity to normal cochlear processing





Separate channels in a 6channel simulation

file=1to6.sfs speaker= token=

Time (s)	0.0.10.20.40.6	0.8	.0 1.2	1.4	1.6
5k 4k - Hz 3k (44dB2k -	=slink(file=d:/talks/cochlear.99/vocoded/1	.sfs,item=1.0	32,start=0,end=	18810,history=s	େଲ୍ୟନ୍ତରେମ୍ବ
1k		- 4- 444 4			
эк 4k - Hz 3k (49dB2k -	TSIINK(file=d:/taiks/cochiear.99/Vocoded/2	.sts,1tem=1.0	22,Start=0,end=	18810,nistory=s	COMPSIDE
1k	Machanakera 038/41/41/2000/06/2010/06/06/2010/06/04		/COLSOND BARRY	14.6.0404466655517411	0
5k 4k - Hz 3k ^{(45dB} 2k - 1k	=slink(file=d:/talks/cochlear.99/vocoded/3	.sfs,item=1.0	32,start=0,end=	18810,history=s	େୟ~େଉଅ
5k	∃ ∃slink(file=d:/talks/cochlear.99/vocoded/4	.sfs,item=1.0	02,start=0,end=	18810,history=s	CG4P3-014
4k - Hz 3k (45dB2k - 1k		Madrice	12.000 (the second	Metalanis	or Office
5k 4k - Hz 3k (35dB2k - 1k	Tslink(file=d:/talks/cochlear.99/vocoded/5	.sfs,item=1.0	32,start=0,end=	18810,history=s	େଇ ମ ୍ୟ ଉତ୍ତି
5k 4k - Hz 3k (35dB2k - 1k	sliver file=d:/talks/cochlear.99/vocoded/6	.sfs	32,start=0,end=	18810,history=s	
Time (s)	¹ 0.0	0.8	.0.1	1	1.6

... and when summed together.









Never mind the quality... feel the intelligibility.









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



Modulation depth matters, too



1000

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.



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)
 - $_{\odot}$ 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