The psychophysics of cochlear implants

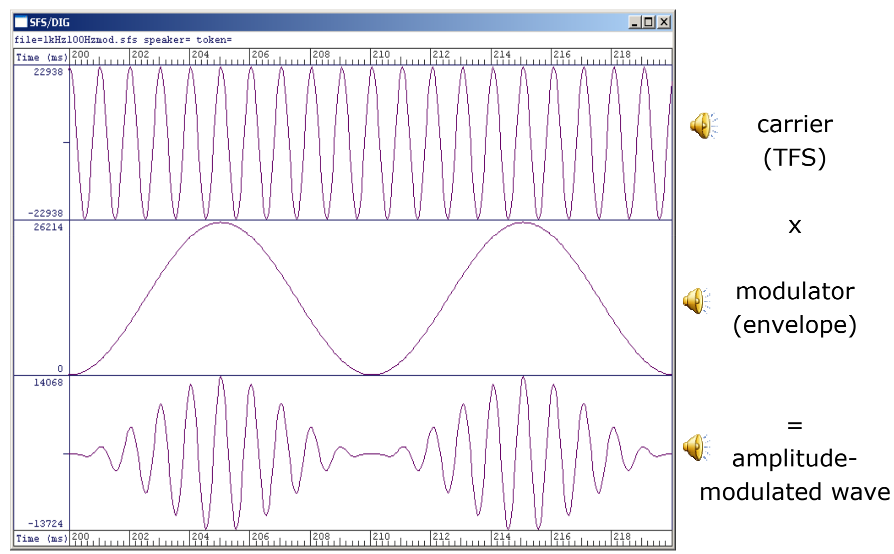
Stuart Rosen
Professor of Speech and Hearing Science
Speech, Hearing and Phonetic Sciences
Division of Psychology & Language Sciences

Prelude
Envelope and temporal fine structure (TFS): What's all the fuss?

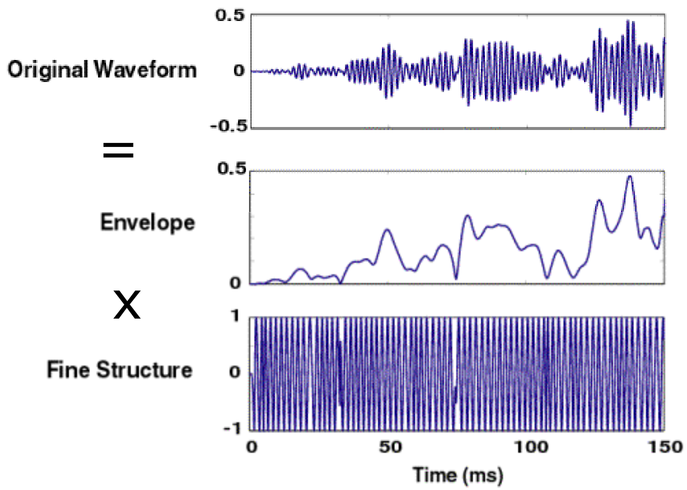
Decomposing waveforms

- Spectral analysis ...
  - Decomposes a wave into a sum of sinusoids to give a *spectrum*
- This particular temporal analysis …
  - Decomposes a wave into the *product* of two (usually) complicated waves known as the *envelope* and the *temporal fine structure* (TFS).

Modulating a wave



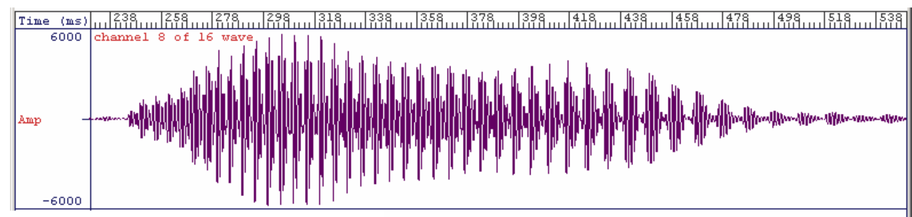
Can work this backwards



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

Fine structure and envelope

- Temporal fine structure – relatively fast – reflects spectral components of sounds in the sound waveform, and periodicity (in some definitions)
- envelope is the slower stuff
- think of all waves as being made by multiplying an envelope against a carrier



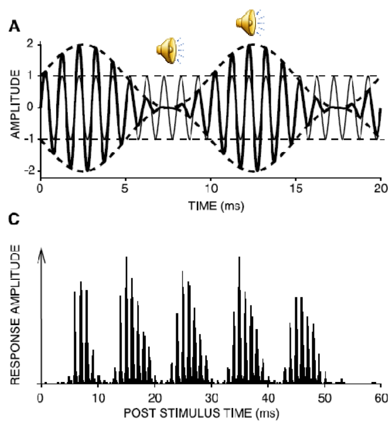
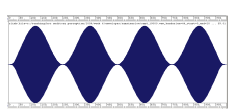


FIG. 1. *A*: superimposed waveforms of an unmodulated 1,000-Hz tone (thin line) and the same tone sinusoidally amplitude modulated (AM) (thick line) at 100% with a modulation frequency of 100 Hz, according to *Equation 1*. Dashed lines indicate the envelope. The amplitude is referenced to the peak amplitude 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 ± 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.

Both (all 3?) kinds of temporal features preserved in the auditory nerve

•Joris *et al.* 2004

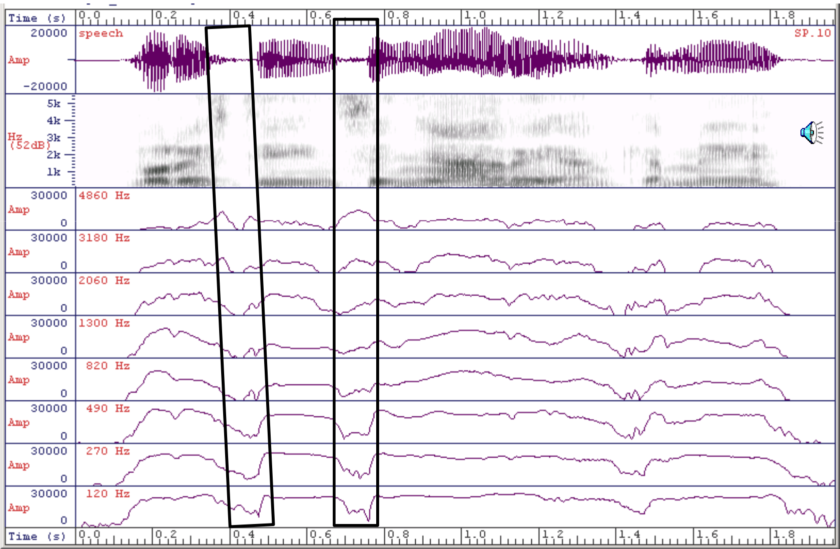
•Slower features too (4 Hz modulations)



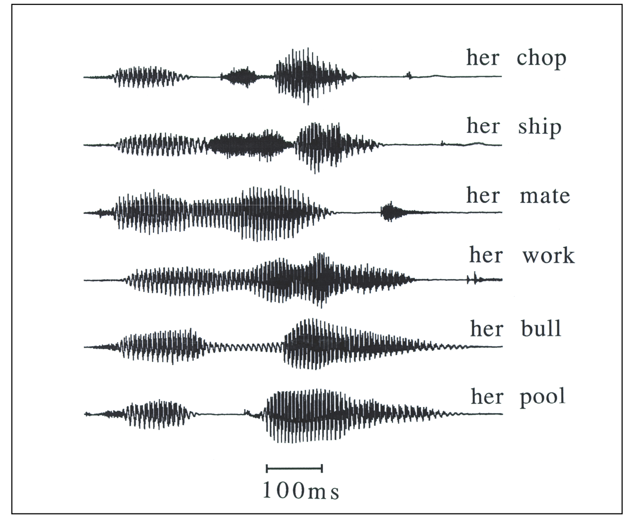
Everyone agrees that …

- '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.

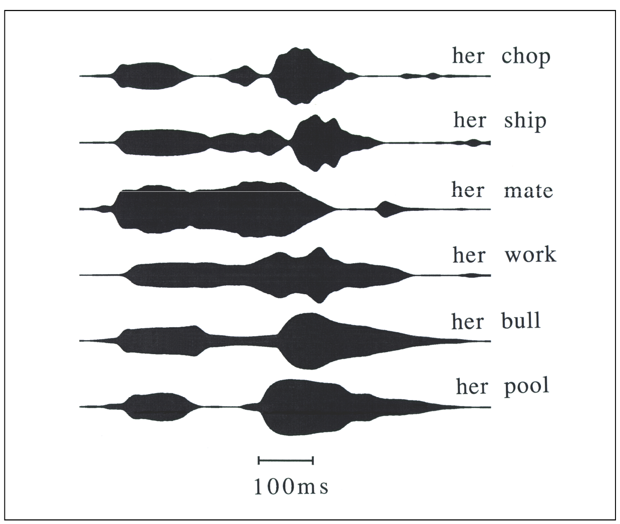
Correlated and uncorrelated (across frequency) envelope modulations



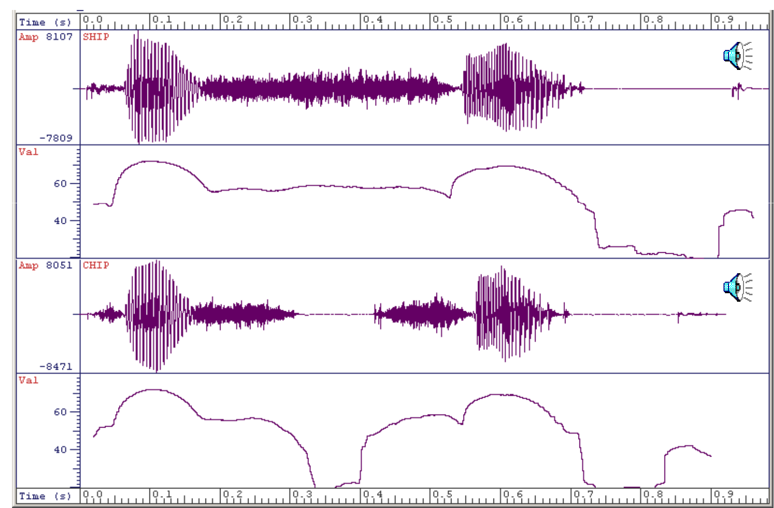
Correlated envelopes in speech – one source of cues to consonants



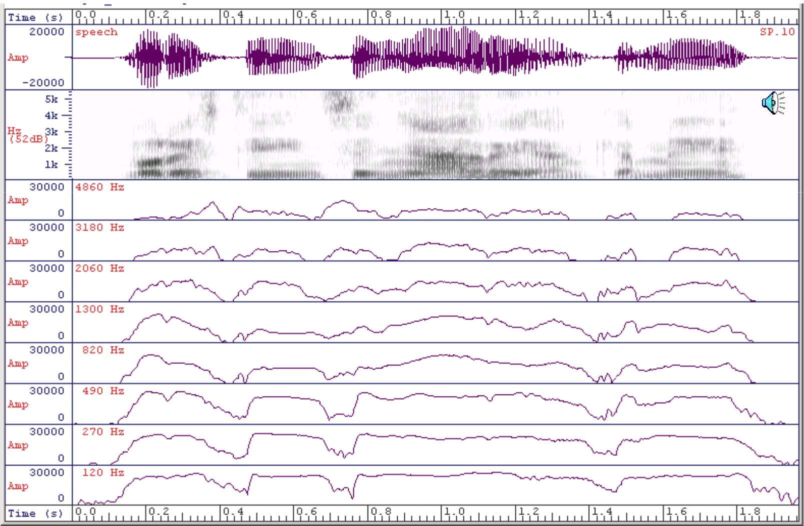
Correlated envelopes in speech – one source of cues to consonants



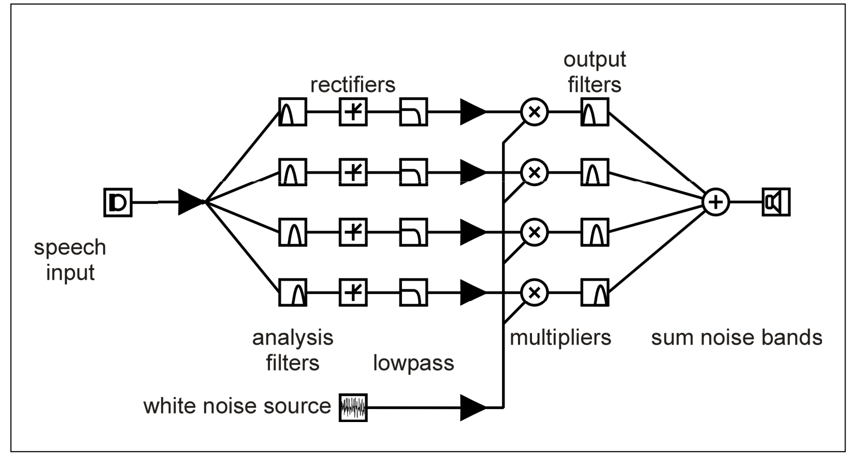
Changing manner of articulation *push ship* vs. *push chip*



Spectral dynamics are encoded in uncorrelated across-channel envelope modulations

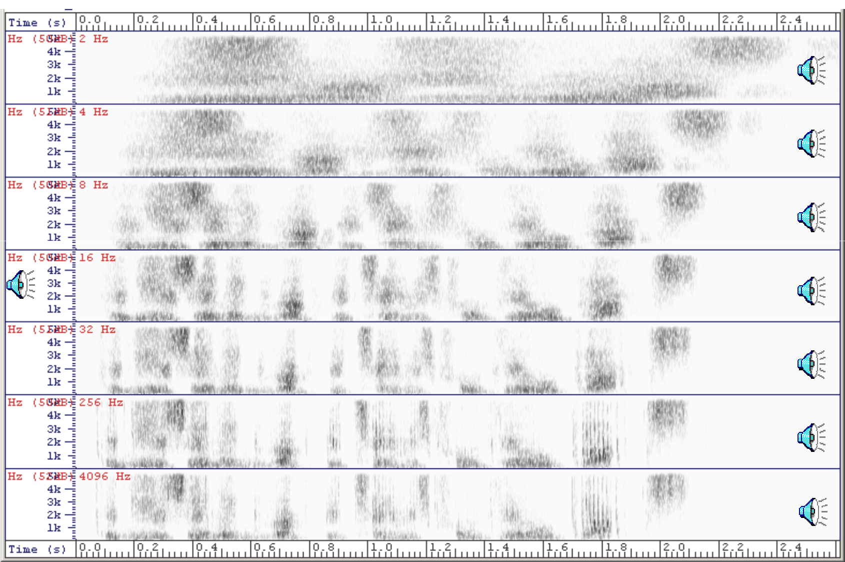


Proof that slow envelopes are sufficient: Noise-excited vocoding



preserves envelopes, destroys TFS

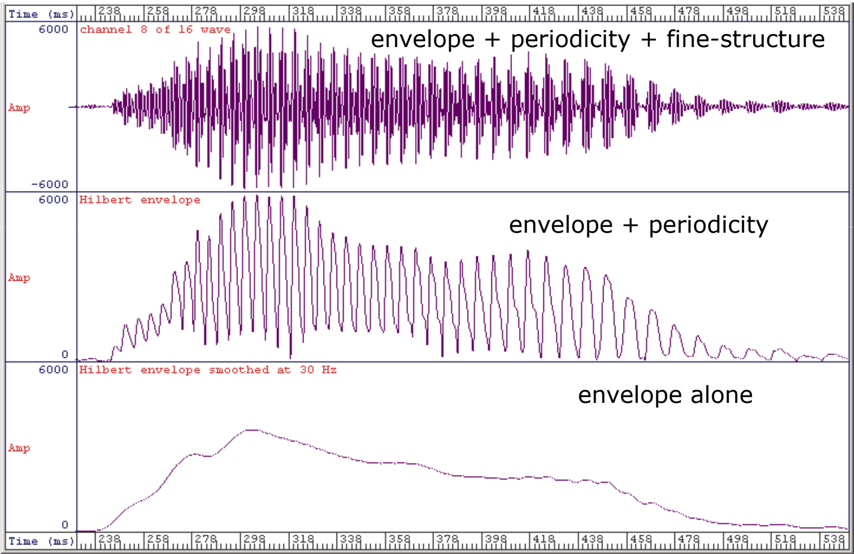
Modulations < ≈10 Hz are most important



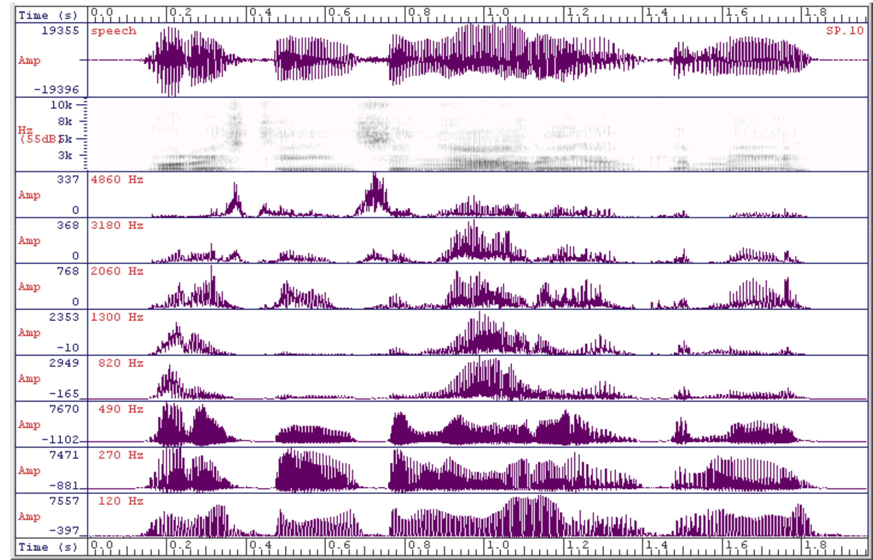
So what’s missing?

- TFS said to be important for …
  - Perception of pitch
    - Intonation and tone
  - ‘Glimpsing’ in noises that vary in level
    - An ability that allows a listener to tolerate higher levels of noises than would otherwise be possible

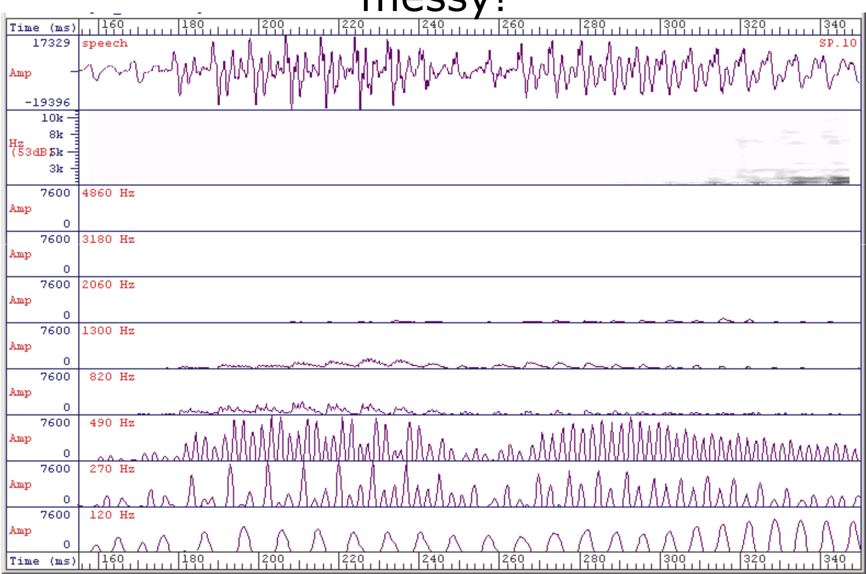
A 3-way partition: typical for NHLs at higher frequencies and CIs



Periodicity in CIs encoded primarily as changes in within channel modulation rate



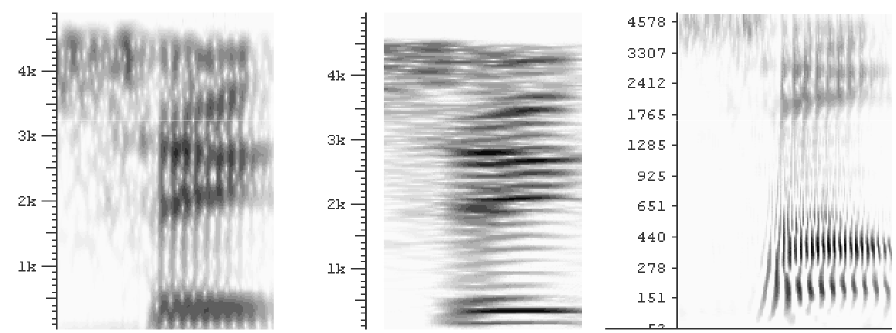
The representation of periodicity can be messy!



NHLs do use TFS for pitch

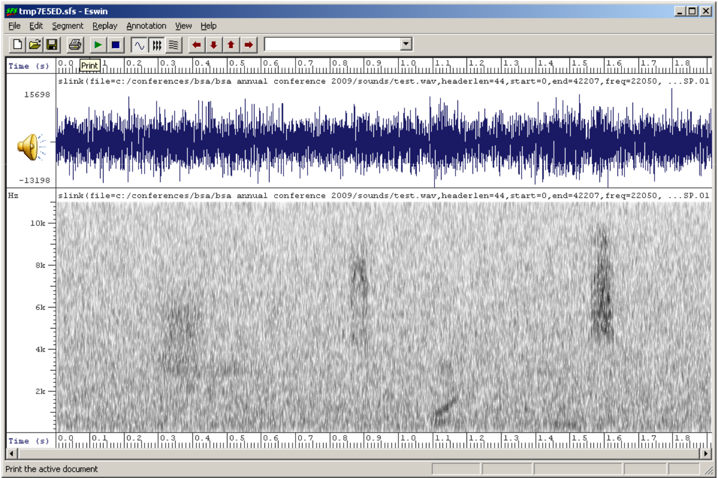
Types of Spectrogram

Wide-band Narrow-band Auditory

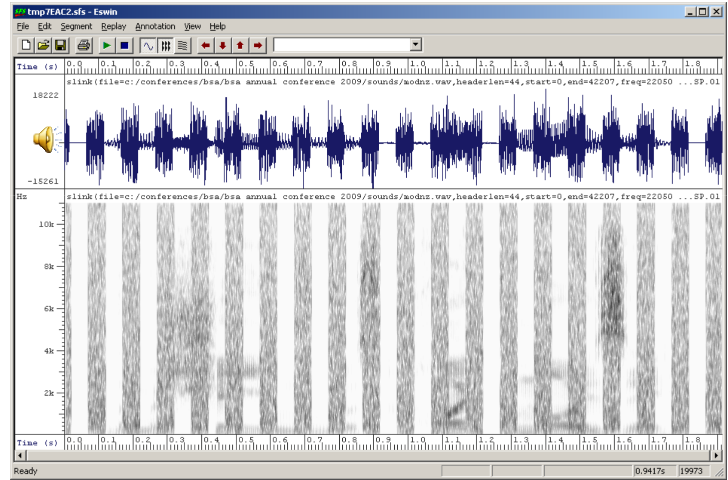


•*An auditory spectrogram looks like a wide-band spectrogram at high frequencies and a narrow-band spectrogram at low frequencies (but with more temporal structure).*

No glimpsing opportunities: A steady-state background noise



But noises are typically not steady …



Does TFS have a role in glimpsing?

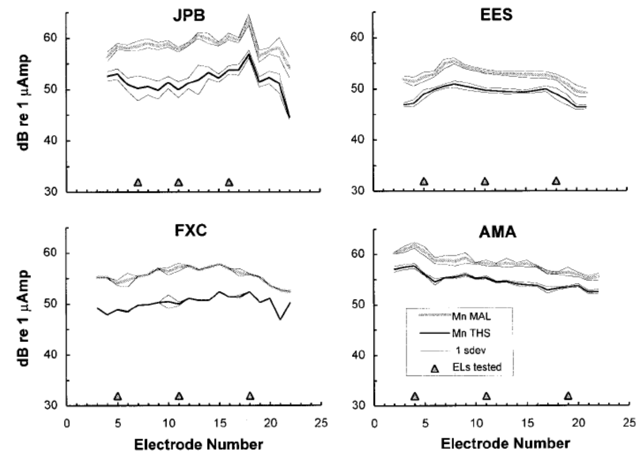
- CI users do not appear to be able to glimpse,
- Nor do NHLs in simulation studies...
- So perhaps TFS (or some aspect of periodicity) is necessary

Summarise

- Waveforms (after any filter bank/spectral analysis) can be decomposed into the product of
  - An envelope (something slow)
    - Or maybe two kinds of envelope
  - A TFS (something fast)
- One limitation of CIs may be poor access to TFS information
  - Also sometimes used as a code word for 'pitch perception' hence necessary for music.

The psychophysics of electrical stimulation in the cochlea

Restricted dynamic range means compression is crucial



Absolute thresholds and maximum acceptable loudness levels
Nelson *et al.* (1996) JASA

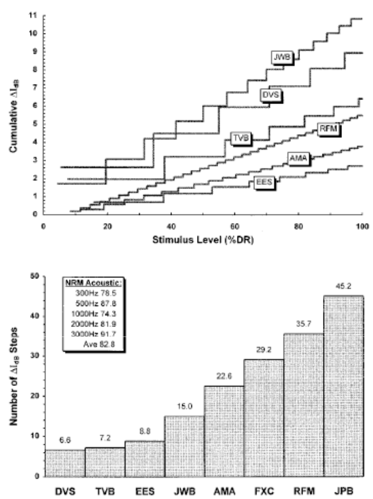


FIG. 9. Cumulative discriminable intensity steps across dynamic range and the number of discriminable intensity steps per subject. *Upper panel*: Cumulative ΔI_{dB} $\{10 \log(I+\Delta I)-10 \log(I)\}$ as a function of stimulus level in percent dynamic range (%DR in dB), which were calculated from the composite Weber functions in Fig. 6. Curves for JPB and FXC were not plotted because they overlapped with the curve for RFM. *Lower panel*: The total number of discriminable intensity steps across dynamic range is given for each of the eight subjects. The total number of discriminable intensity steps for normal acoustic hearing, calculated from Weber fractions reported by Schroder *et al.* (1994), are shown for each of five frequencies within the inset.

Intensity jnds in electrical (opposed to acoustic) stimulation:

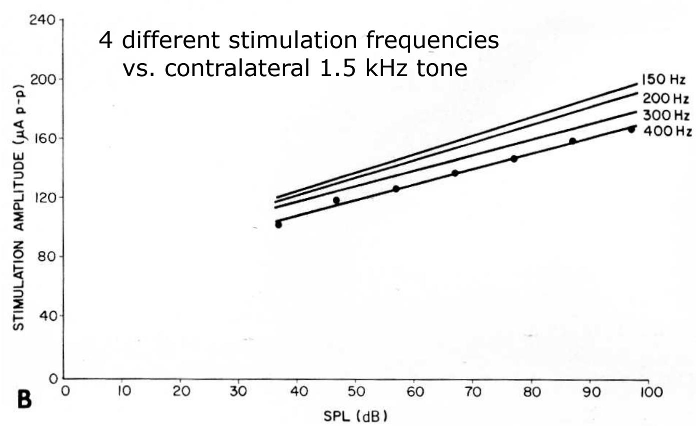
1) 'miss' Weber's Law more

2) are typically smaller, but not by enough to offset reduced dynamic range.

CI users here had 7-45 discriminable steps in the total dynamic range, compared to ≈ 83 in normal hearing

Nelson *et al.* (1996) JASA

Acoustic/electrical loudness matches



Eddington *et al.* 1978 Ann Otol Rhinol Laryngol

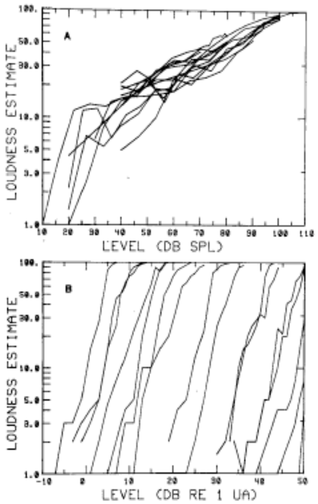
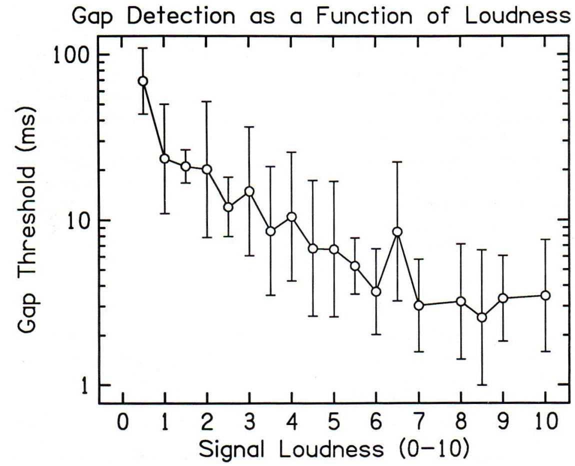


Fig. 3. Comparison of loudness vs. stimulus intensity curves for loudness estimates from normals (A) and implant subjects (B). The loudness estimation data for the four normal hearing subjects was all collected at 1000 Hz. The exponent of the power function was inversely related to the dynamic range for electrical stimulation. Examples shown are for 100 Hz, where the dynamic range was 30 dB, and for 1000 Hz, where the dynamic range was only 18 dB.

Loudness grows much faster in electrical stimulation (hyper-recruitment!)

Temporal resolution: gap detection



Shannon 1993

Temporal resolution: modulation detection (100 Hz)

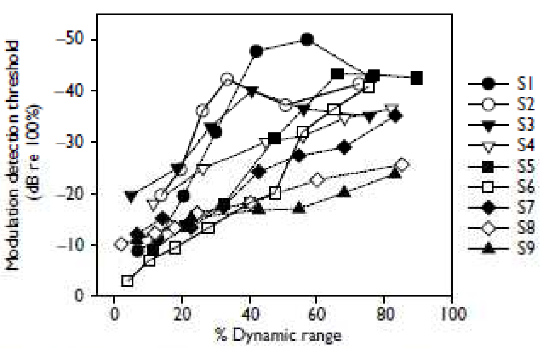
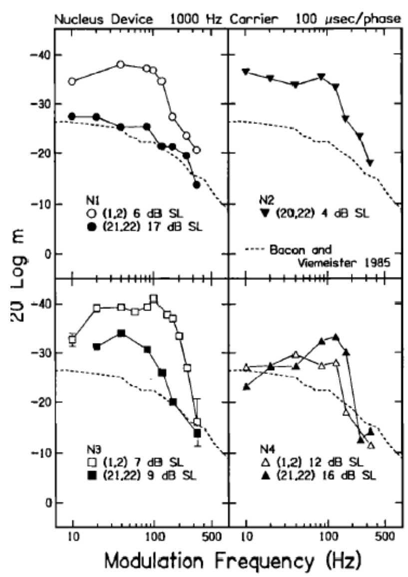


Fig. 1. Modulation detection thresholds as a function of the percentage of subjects' electric dynamic range

More dependent on level (as for intensity jnd's)

Fu 2002 NeuroReport



Shannon 1992 J Acoust Soc Amer

Temporal resolution: TMTFs

More dependent on level

Otherwise similar to normal listeners (dashed lines)

Relationships to performance with speech

modulation detection thresholds measured at 100 Hz, at a number of levels (previous slide)

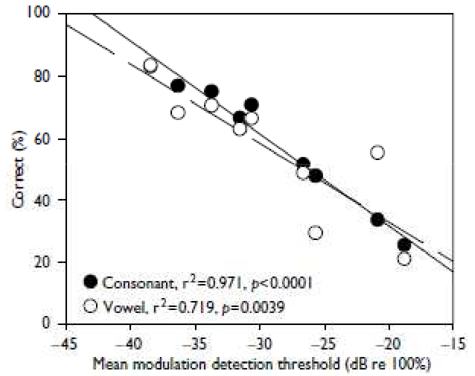


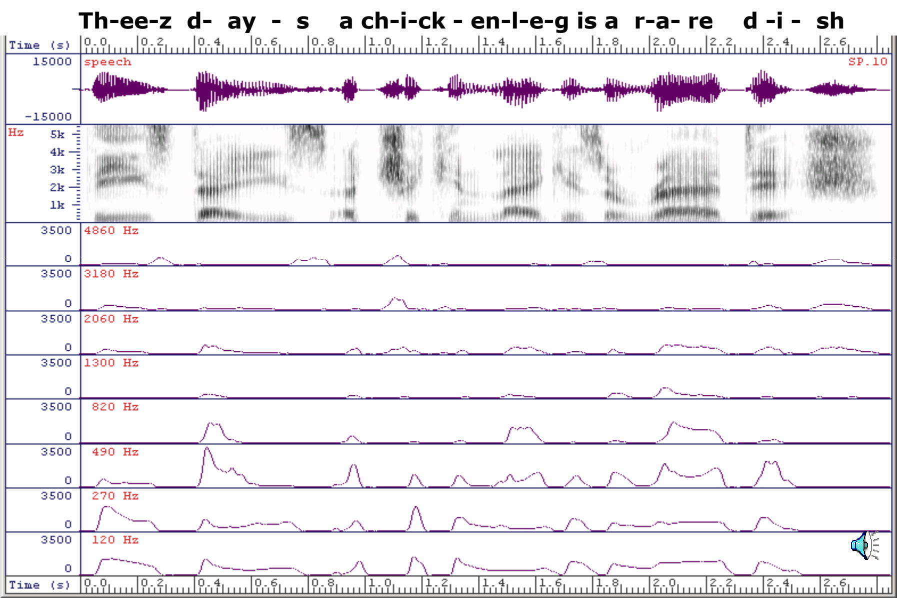
Fig. 2. Correlation between phoneme identification (percent correct) and subjects' mean modulation detection thresholds (calculated across each subject's entire dynamic range). Consonant scores and linear regression are shown by the filled circles and solid line. Vowel scores and linear regression are shown by the open circles and dashed line.

Fu 2002 NeuroReport

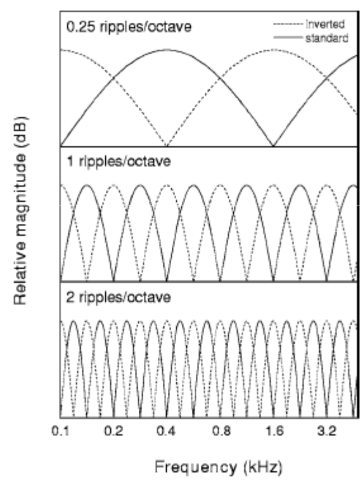
Perceiving variations in amount of activity across electrodes

- Essential for signaling of …
  - spectral shape
- Spectral shape is encoded by relatively slow level changes across electrodes
- Recall
  - preservation of fast modulation rates not necessary for intelligibility in noise-vocoded speech

Slow level changes across channels



Discrimination of rippled noise



find the maximum ripple density at which it is possible to discriminate 'standard' ripple noise from its inverted version

'This test is hypothesized to provide a direct measure of the ability of listeners to perceive the frequency locations of spectral peaks in a broadband acoustic signal.'

Henry *et al.* 2005 J Acoust Soc Am

Discrimination of rippled noise

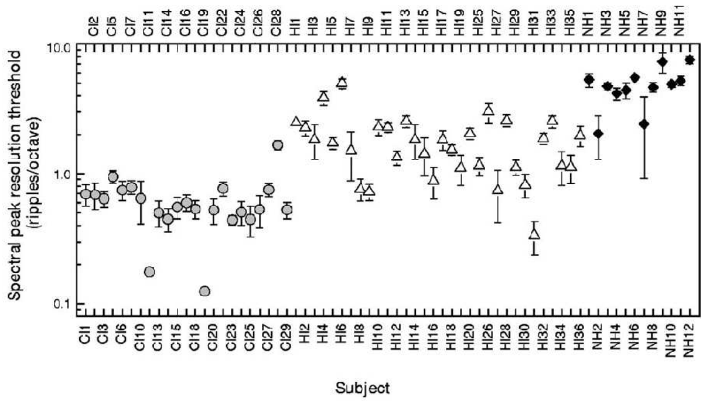


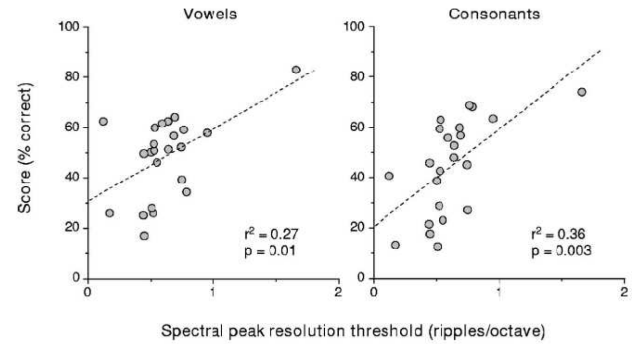
FIG. 2. Thresholds for spectral peak resolution for NH, HI, and CI subjects. Error bars represent ± one standard deviation.

Henry *et al.* 2005 J Acoust Soc Am

Relationships to performance with speech in quiet

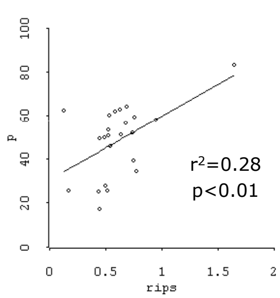
12 hVd by 20 talkers

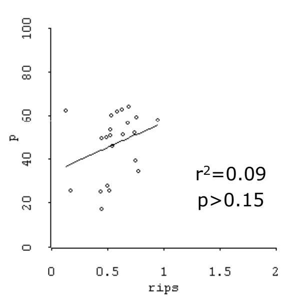
16 VCVs by 4 talkers



Henry *et al.* 2005 J Acoust Soc Am

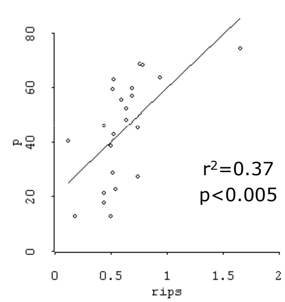
Statistical interlude: The effect of outliers

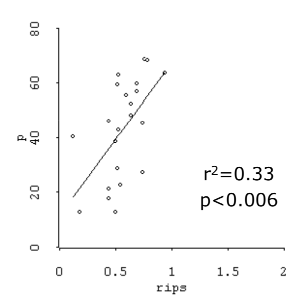




vowels

Statistical interlude: The effect of outliers





consonants

Relationships to performance with speech in noise

SRT determined for selection of one of 12 spondees

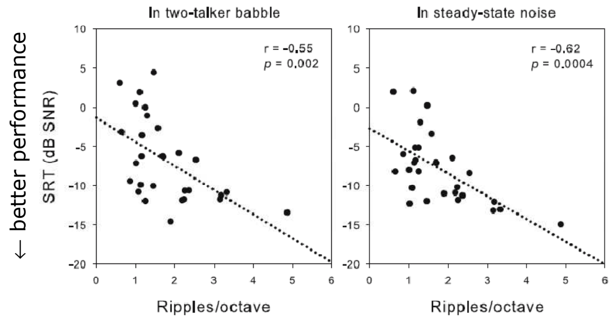


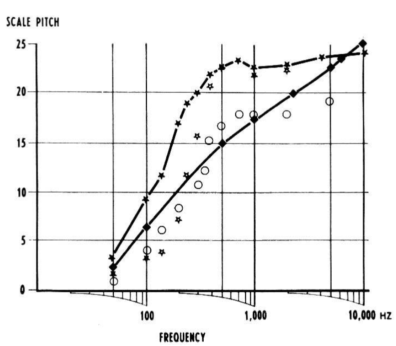
FIG. 6. Spectral-ripple discrimination is correlated with speech perception in noise. The figure shows the relationship between the spectral-ripple thresholds and SRTs in two-talker babble (*left panel*) and steady-state noise (*right panel*) using data from the first six repetitions. Linear regressions are represented by the *dotted lines*.

Won *et al.* 2005 JARO

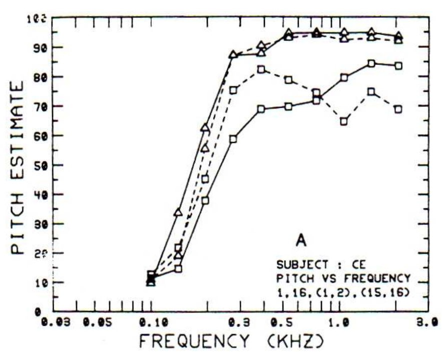
Why is speech melody (*voice pitch)* important to hear?

- Contributes to speech intelligibility in all languages
- A good supplement to lipread information
- May play an important role in separating speech from background noises
- Appears to play a more crucial role for the young child developing language
- Crucial in so-called *tone* languages

Pitch based on a purely temporal code



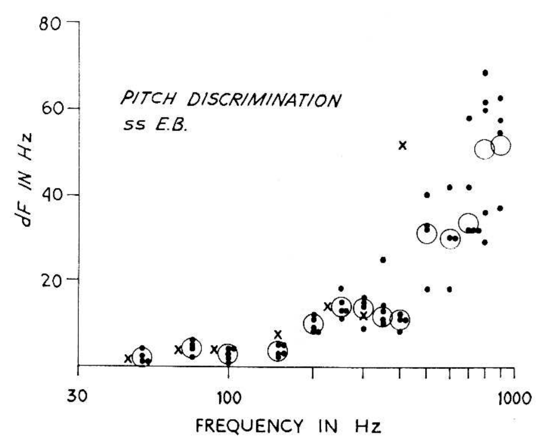
Merzenich *et al.* 1973



Shannon 1993

limited to 300 Hz or so

Pitch based on a purely temporal code



Best normal performance for normal listeners about 0.2 % over entire range

Merzenich *et al.* 1973

Melody recognition

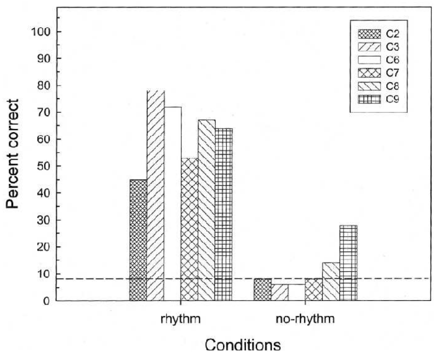
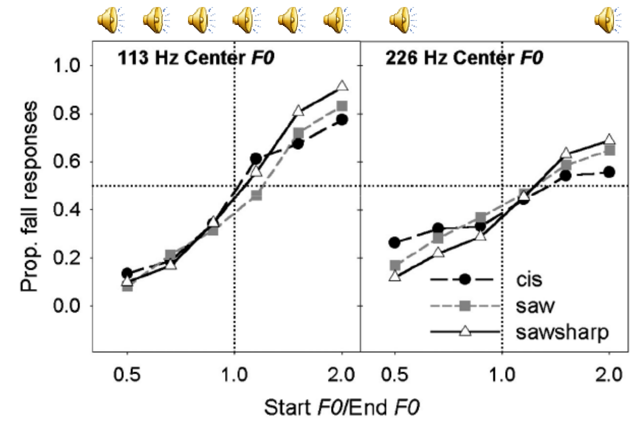


Figure 4. Melody identification scores from individual cochlear implant listeners with the original melodies. The horizontal dashed line indicates the mean chance performance. The vertical bars represent different subjects in each condition.

12 songs familiar to most people, synthesised with and without natural rhythm

Kong *et al.* (2004)

CI users classifying rise/fall contours on diphthongs



Green *et al.* 2004 J Acoust Soc Amer

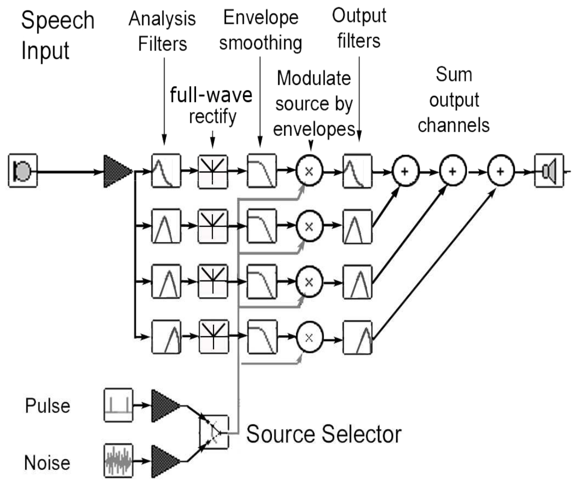
Tones in Mandarin Chinese

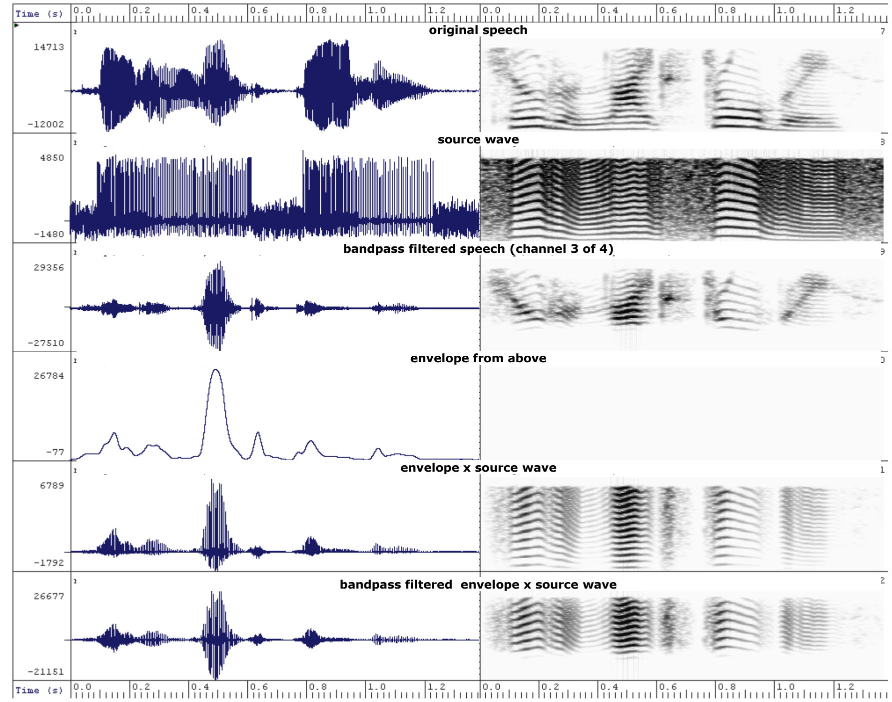
	STANDARD CHINESE **ma**		
	Chinese Character	Tone symbol	Tone description
mother	媽	˥	high level
hemp	麻	˧˥	high rising
horse	馬	˨˩˦	low falling
scold	罵	˥˩	high falling

How important is the loss of voice pitch to understanding speech in quiet?

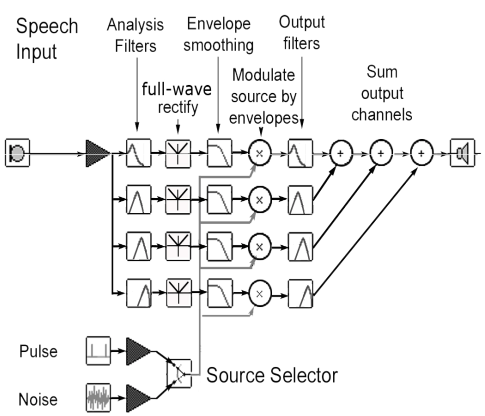
- Eliminating tonal contrasts from speech still leaves tone languages intelligible …
- because no single acoustic feature is indispensable in any language.
- Here, we trade off spectral resolution against presence or absence of tone (voice pitch variations/speech melody).

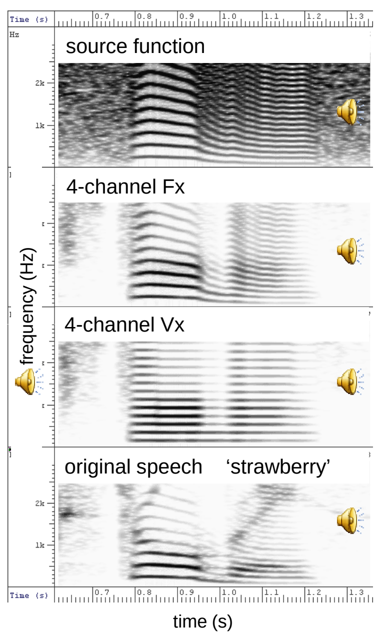
Vocoding as a way to trade off spectral resolution and tone





4-channel vocoding





Experimental design

- Three tone languages
  - Mandarin, Cantonese & Yoruba
- Two non-tonal languages
  - English & Lithuanian
- Presented to groups of native listeners varying …
  - numbers of channels (1, 2, 4, 8 and 16)
  - presence or absence of informative tone

Examples of the stimuli

VxNx FxNx

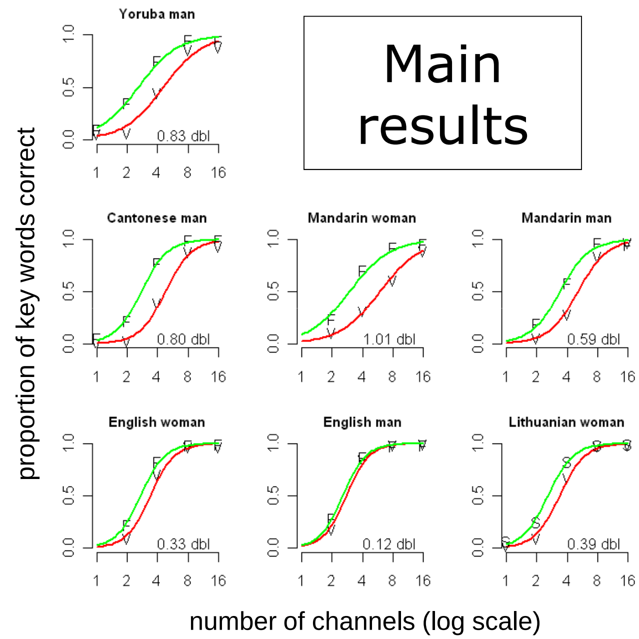
1

2

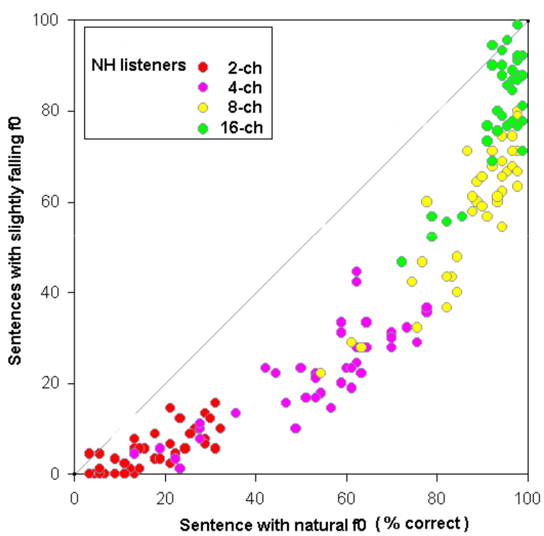
4

8

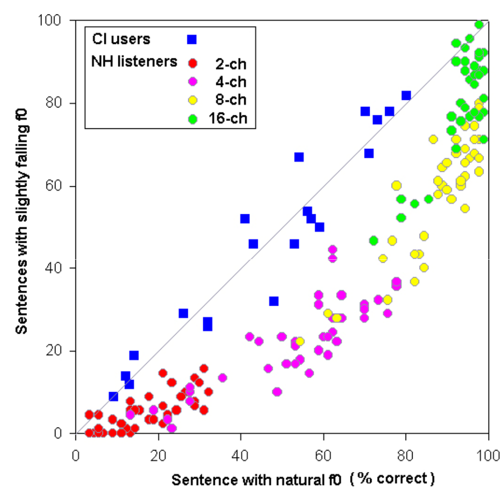
16

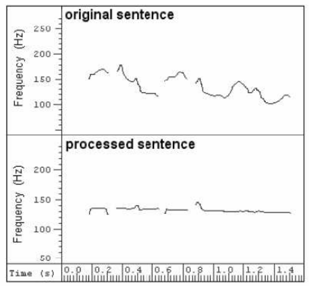


Results for Mandarin (simulations)



No use of linguistic tone in Mandarin by CI users





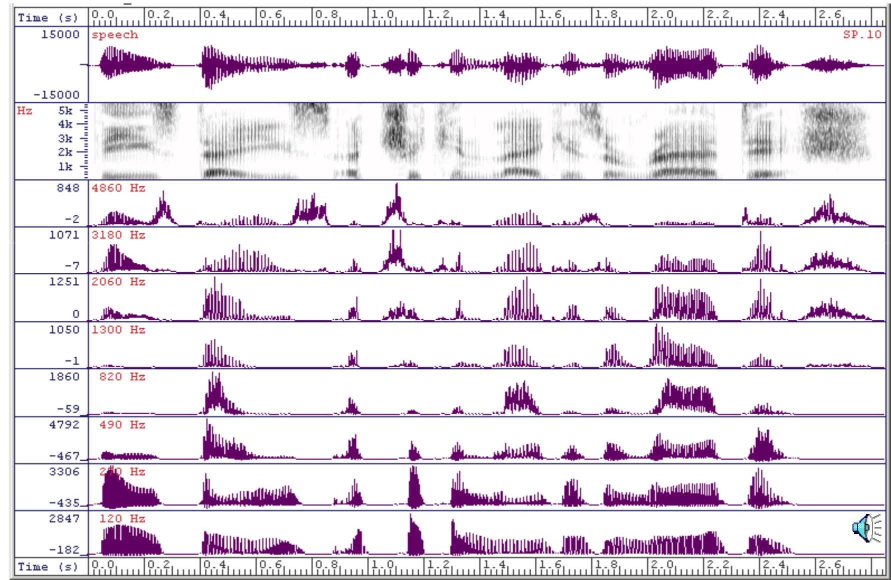
from the PhD thesis of Yu-Ching Kuo, 2006

Conclusions

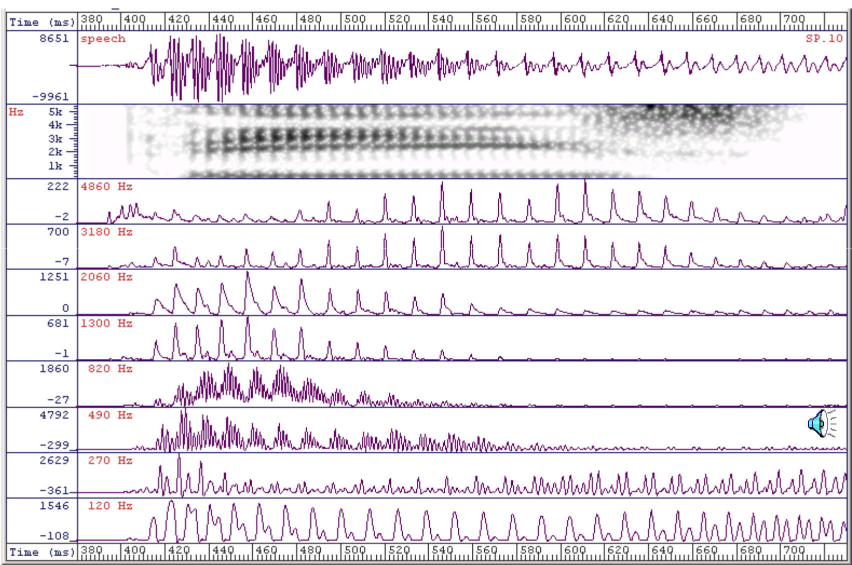
- Variations in fundamental frequency contribute to intelligibility in all languages …
- but they are considerably more important in tone languages
- Getting tone into cochlear implants could be worth as much as a doubling in the number of channels.

Melody coded as periodicity in rapid within-channel patterns

Th-ee-z d- ay - s a ch-i-ck - en-l-e-g is a r-a- re d -i - sh



The representation of melody can be messy!



Perception of fundamental pitch in complex waves is very poor

- Lower harmonics cannot be resolved as in normal hearing
- Phase-locking seems ‘different’
- Mis-match between place of excitation and temporal pattern may be important