Who are cochlear implants for?

- People with little or no hearing
  - and little conductive component to the loss
- who receive little or no benefit from a hearing aid.
- Implants seem to work best in …
  - adults who had a significant period of relatively good hearing before becoming profoundly deaf, and who developed good language.
  - children who are young enough to develop language through an implant.

Essential feature

- substitute for faulty or missing inner hair cell …
- by direct electrical stimulation of residual auditory nerve fibres
  - but brain stem implants are also being used
- Need, at a minimum …
  - microphone + ‘processor’
  - electrodes in the cochlea
  - a way to connect them (radio transmission)

1. Sound is received by the microphone of the speech processor.

2. The sound is digitized, analyzed and transformed into coded signals.

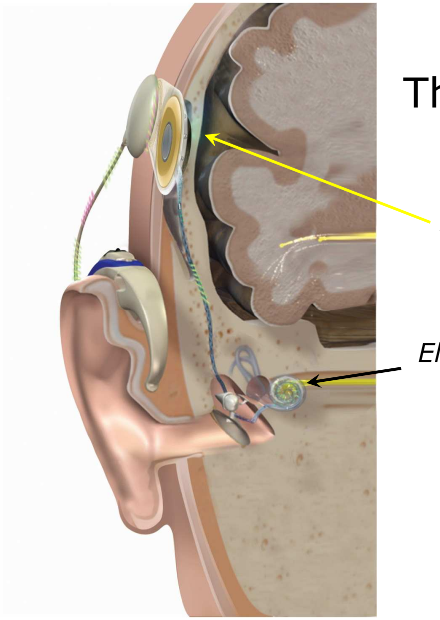
3. Coded signals are sent to the transmitter.

4. The transmitter sends the code across the skin to the internal implant where it is converted to electric signals.

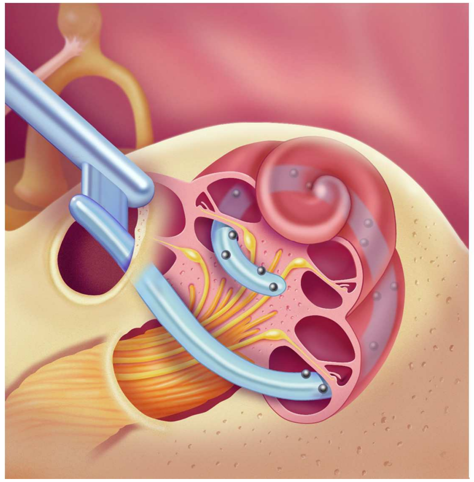
5. Electric signals are sent to the electrode array to stimulate the residual auditory nerve fibres in the cochlea.

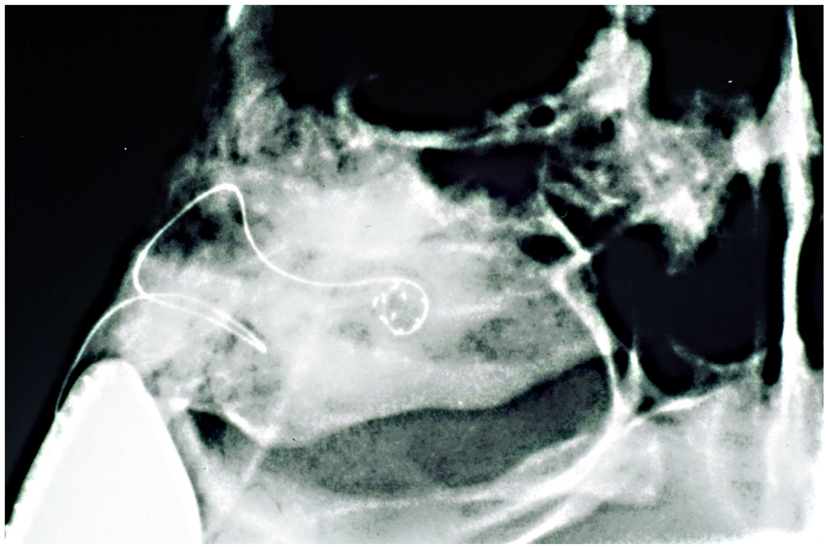
6. Signals travel to the brain, carrying information about sound.

The implant in place



The electrode array





What are the *essential* purposes of a speech processor?

- To transduce acoustical signals into an electrical form.
- To process the acoustic signal in various ways (*e.g.*, filter, compress).
- To convert (or code) the resulting electrical signals into a form appropriate for stimulation of the auditory nerve.

What other functions can and might be implemented in a speech processor?

- Minimising the effects of background noise.
- The possibility of different processing schemes for different situations.
- Enhancing speech features that contribute most to speech intelligibility.

What should an implant do?

- Mimic the most important functions of the normal ear.
- So what does a normal ear do?
  - frequency analysis
  - amplitude compression
  - preservation of temporal features, bot slow and fast (e.g., through envelope following and phase locking)

Common elements in speech processing

- A microphone to transduce acoustic signals into electrical ones.
- Amplitude compression to address the very limited dynamic range of electro-cochlear stimulation.
- Use of the 'place' principle for multiple electrodes (mapping low to high frequency components onto apical to basal cochlear places).

But speech processing schemes vary significantly in other ways

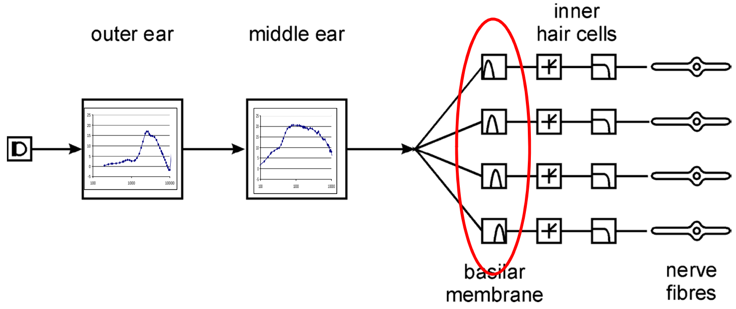
- Pulsatile vs. continuously varying ('wavey') stimulation.
  - Not to be confused with analogue vs. digital implementations. All electrical stimulation is analogue.
- Simultaneous vs. non-simultaneous presentation of currents to different electrodes.
  - Non-simultaneous stimulation *requires* pulsatile stimulation

Multi-channel systems

- All contemporary systems present different waveforms to different electrodes
  - to mimic the frequency analysis of the normal mammalian cochlea.
- Think of the peripheral auditory system as analogous to a *filter bank*.

The filter bank analogy

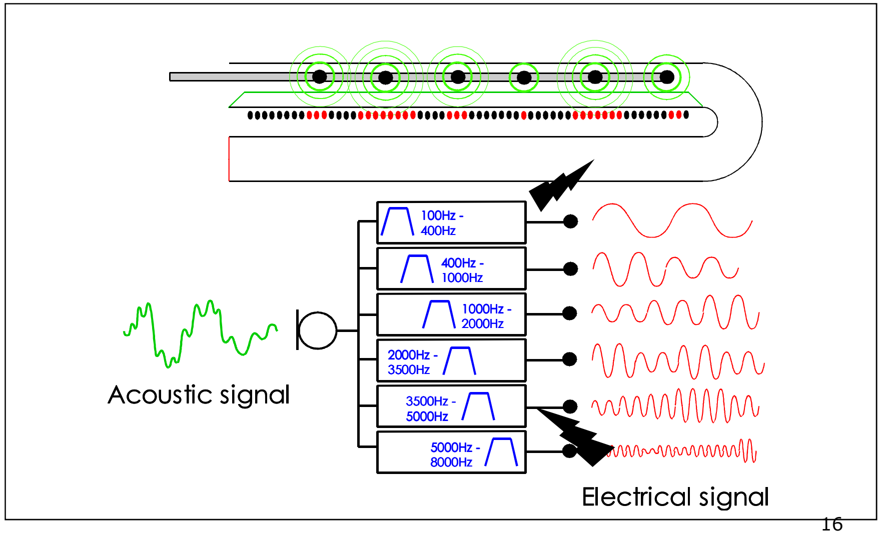
- Imagine each afferent auditory nerve fibre has a bandpass filter attached to its input
  - centre frequencies decreasing from base to apex



The no-brainer cochlear implant speech processing strategy …

- Use an electronic filter bank to substitute for the auditory filter bank (the mechanics of the basilar membrane).

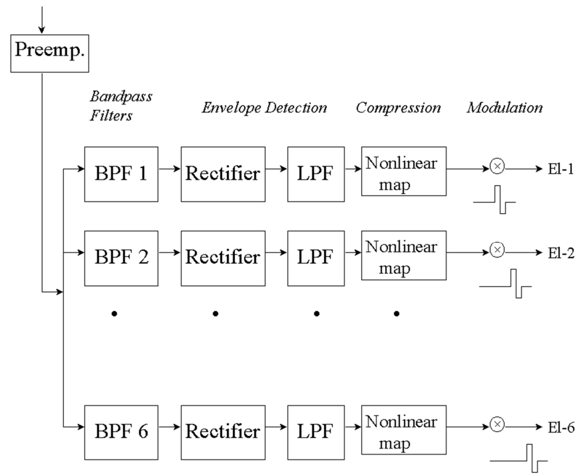
A simple speech processing scheme for a cochlear implant: Compressed Analogue (CA)



The most common current method: Continuous Interleaved Sampling (CIS)

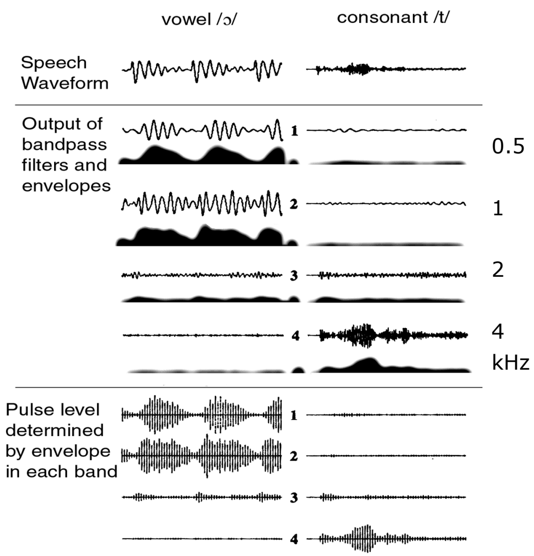
- Use a filter bank approach to represent spectral shape …
- with non-simultaneous pulatile stimulation to minimise electrode interactions
- with pulse amplitudes modulated by the *envelope* of the bandpass filter outputs.

Continuous Interleaved Sampling

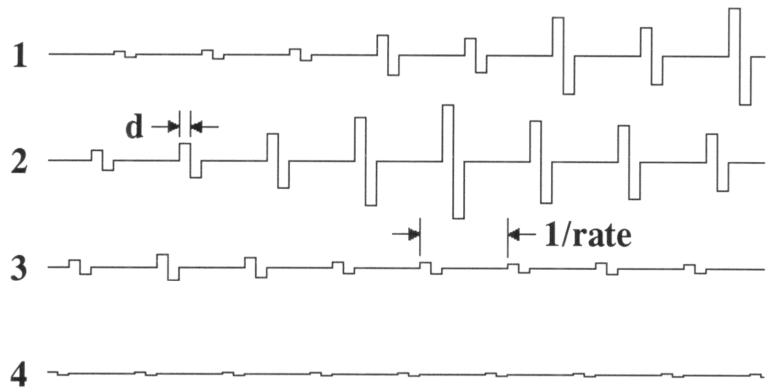


from Philipos Louizou: http://www.utdallas.edu/~loizou/cimplants/tutorial/

Continuous Interleaved Sampling



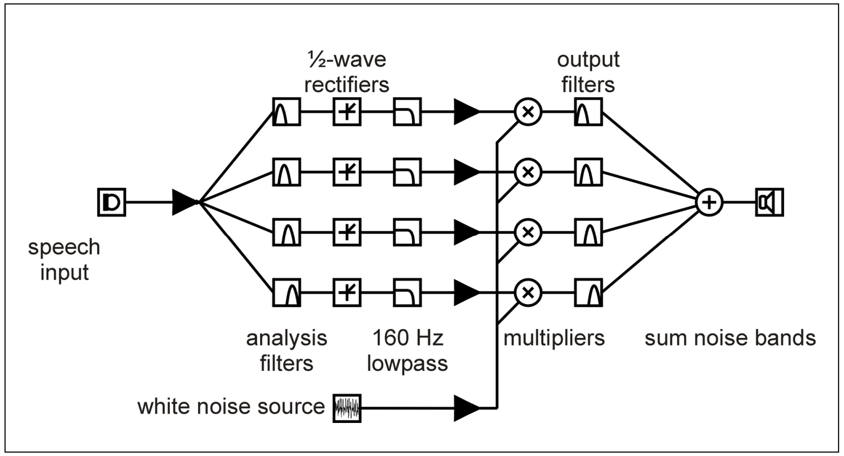
CIS in detail



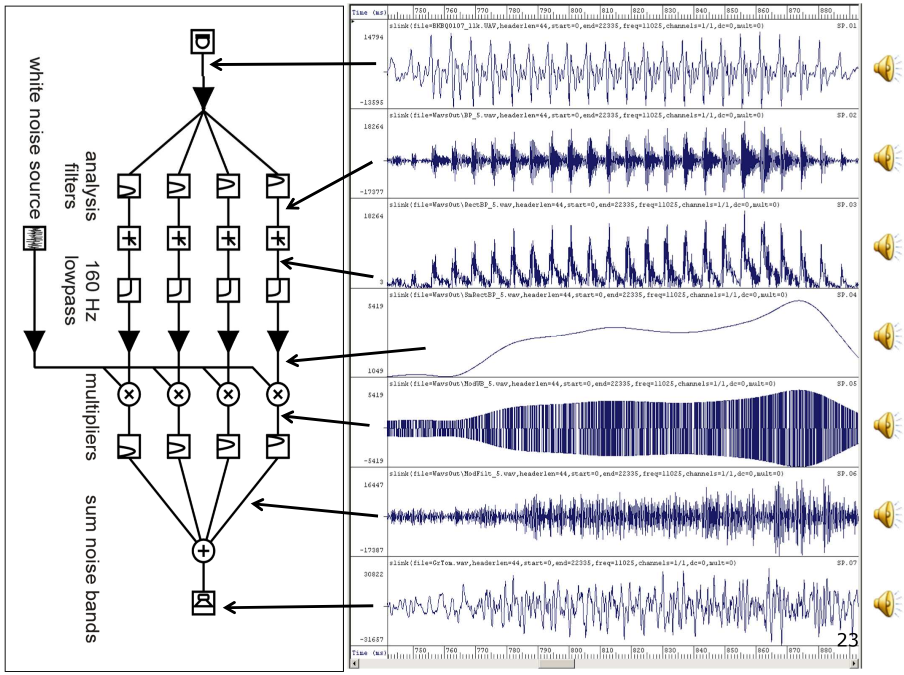
Simulations can give us some idea of what an implant user might experience
But ...*caveat perceptor*!

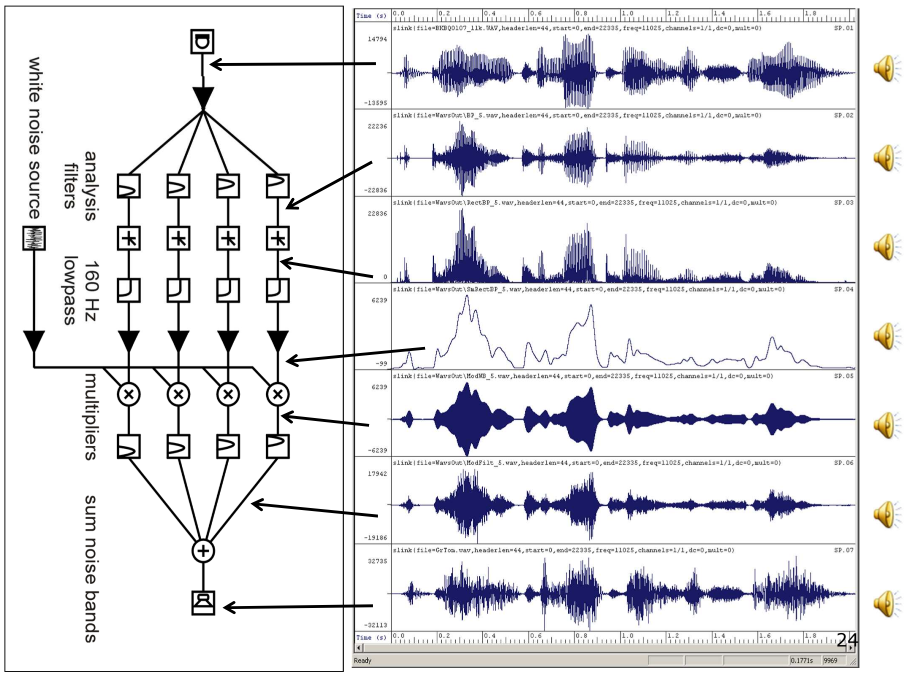
- These are not exactly what an implant sounds like …
- but you can get some idea about the kind of information that gets through.

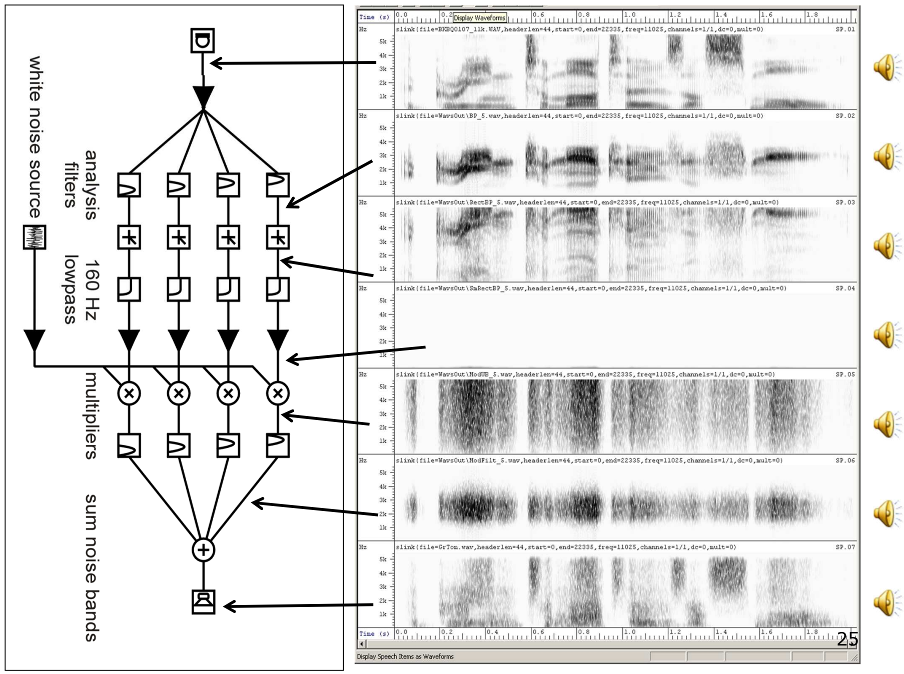
Noise-excited Vocoding



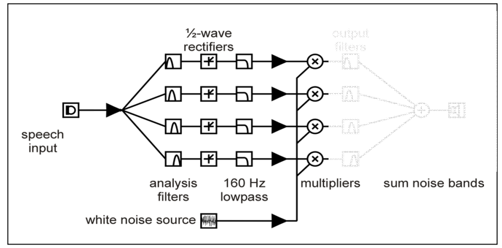
Note important variants in rectification, lowpass filter cutoffs, etc.

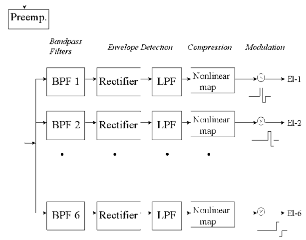


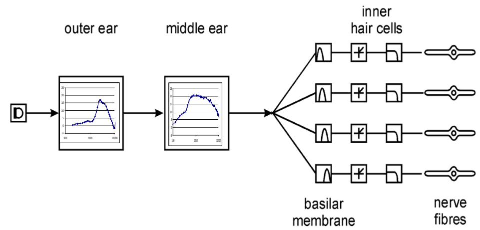




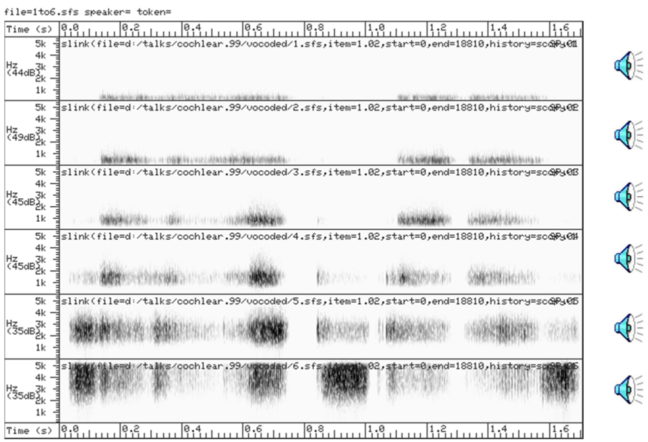
Note similarity to CIS (and normal cochlear) processing





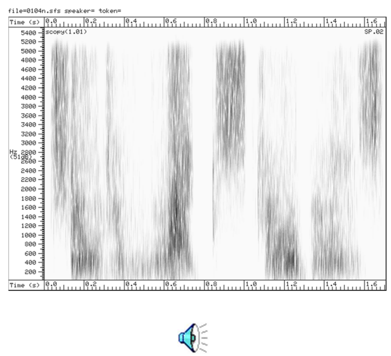


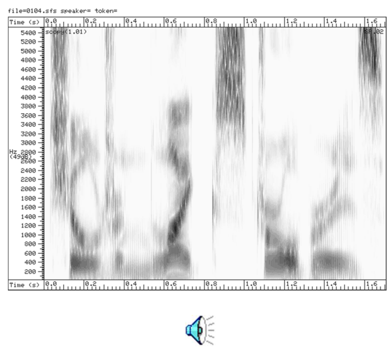
Separate channels in a 6-channel simulation



... and when summed together.

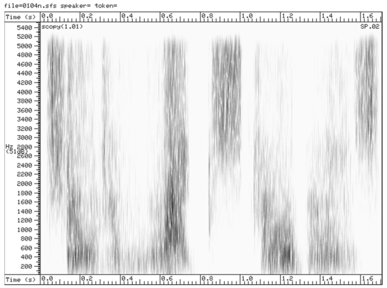
Children like strawberries.

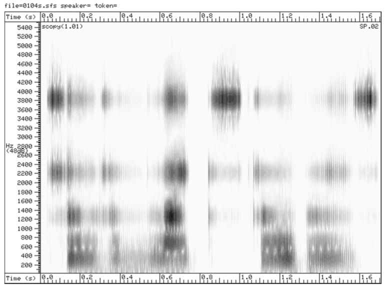




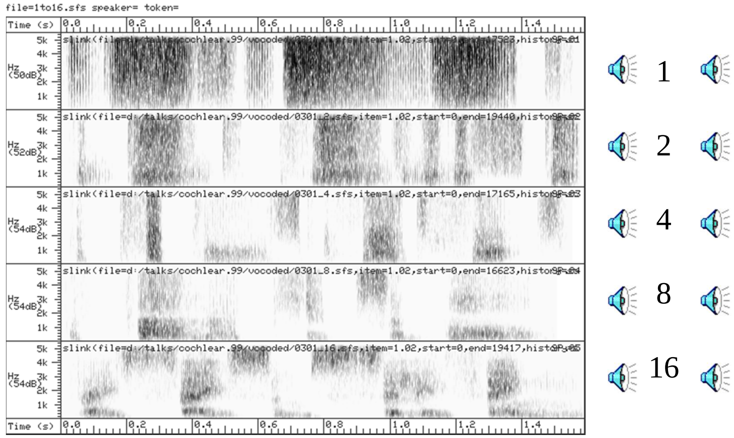
Never mind the quality...
feel the intelligibility.

Children like strawberries.





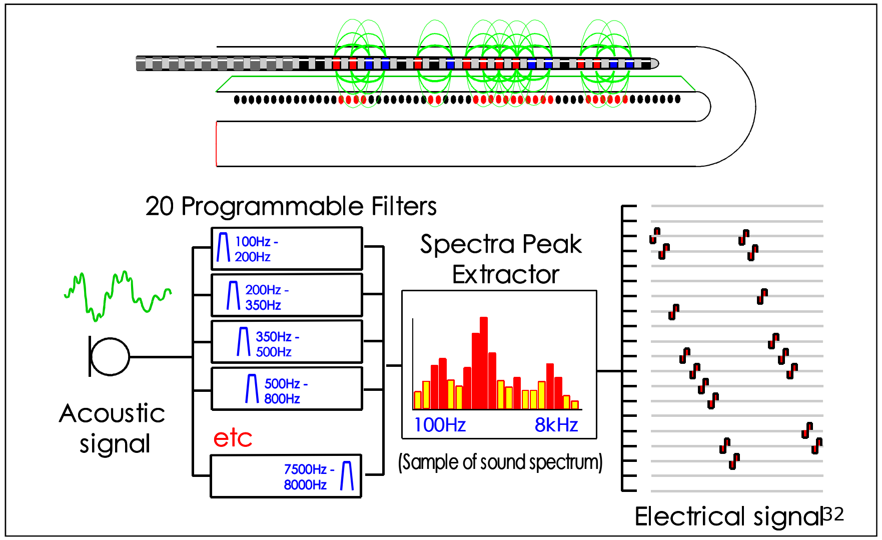
Effects of channel number



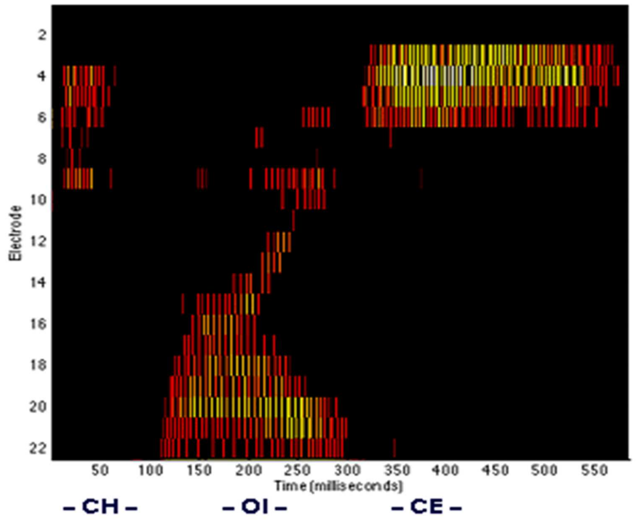
Other schemes:
Necessity is the mother of invention

- The problem (historically)
  - How could devices running at relatively slow rates be used for CIS, which required high rates of pulsatile stimulation?
- The solution
  - Pick and present pulses only at the significant peaks in the spectrum.

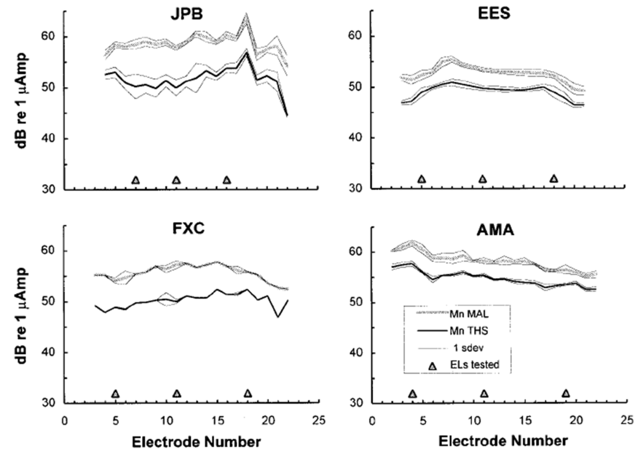
Spectral Peak Strategy – SPEAK
(n of m strategies)



SPEAK stimulation pattern



Restricted dynamic range means compression is crucial



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

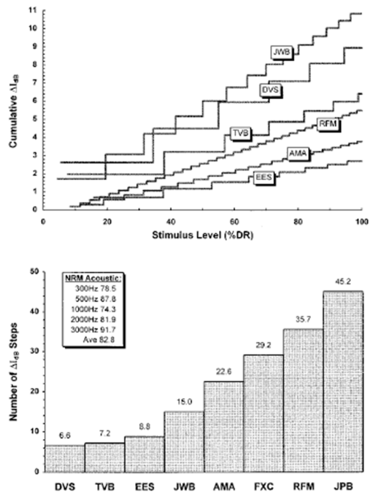


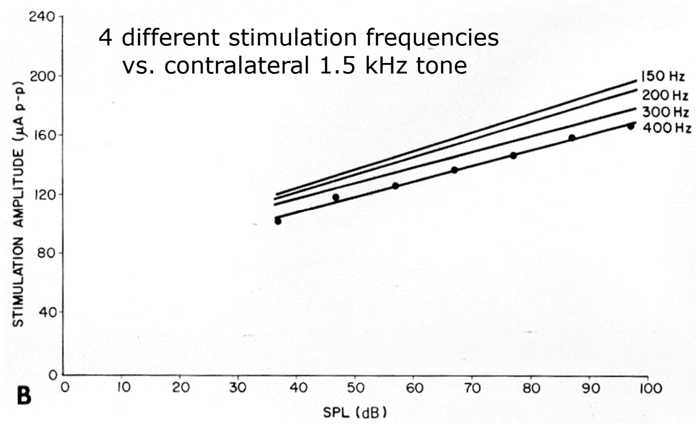
FIG. 9. Cumulative discriminable intensity steps across dynamic range and the number of discriminable intensity steps per subject. *Upper panel*: Cumulative ΔI_{dB} $\{10\log(I+\Delta I)-10\log(I)\}$ as a function of stimulus level in percent dynamic range (%DR in dB), which were calculated from the 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

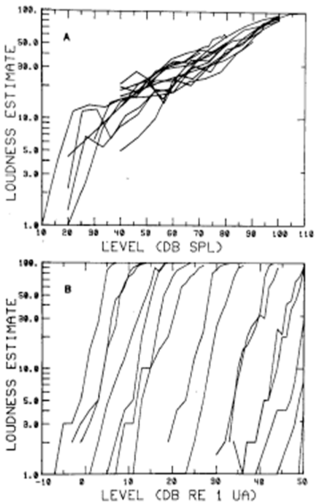
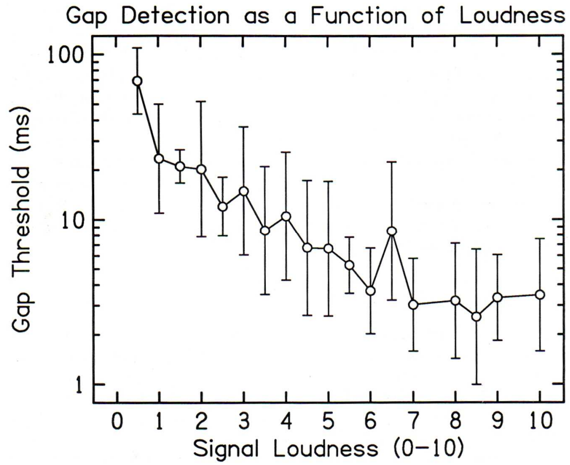


Fig. 3. Comparison of loudness vs. stimulus intensity curves for loudness estimates from normals (A) and implant subjects (B). The loudness estimation data for the four normal hearing subjects was all collected at 1000 Hz. The exponent of the power function was inversely related to the dynamic range 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)

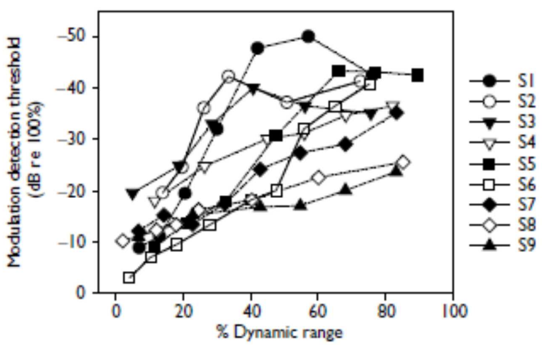
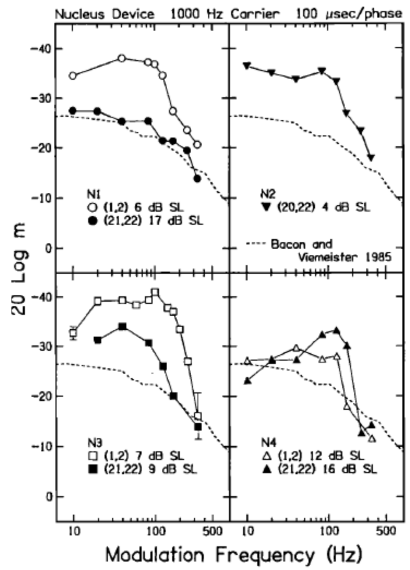


Fig. I. Modulation detection thresholds as a function of the percentage of subjects' electric dynamic range

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

Fu 2002 NeuroReport



Temporal resolution: TMTFs

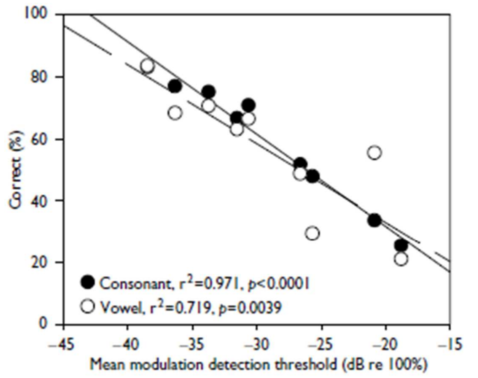
More dependent on level

Otherwise similar to normal listeners (dashed lines)

Shannon 1992 J Acoust Soc Amer

Relationships to performance with speech

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



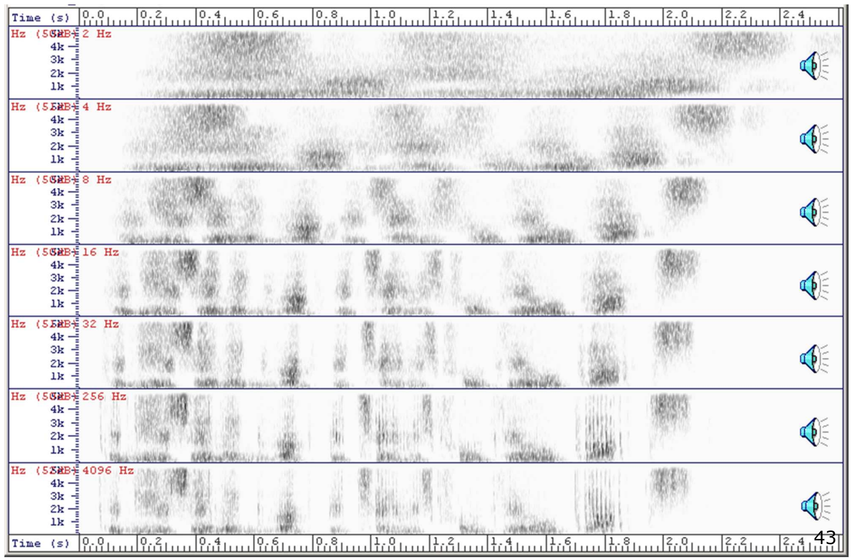
Fu 2002 NeuroReport

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.

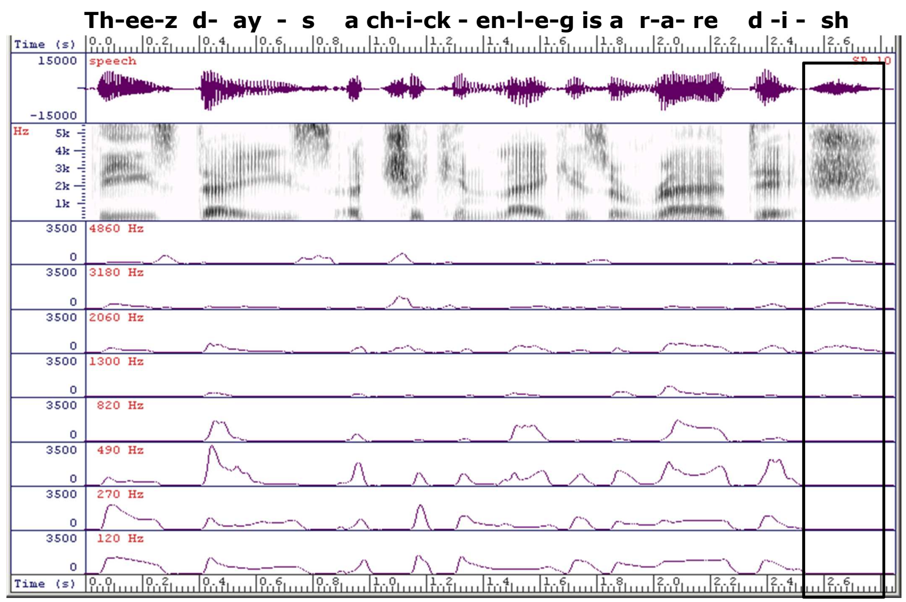
Perceiving variations in amount of activity across electrodes

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

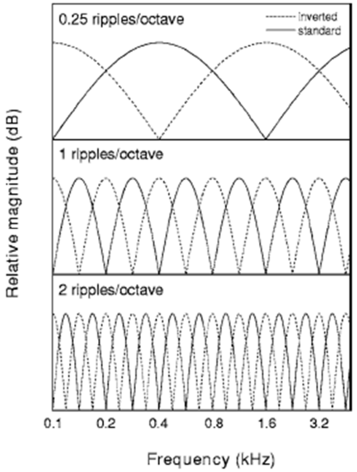
Restricting modulation rates allowable in noise-excited vocoding



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

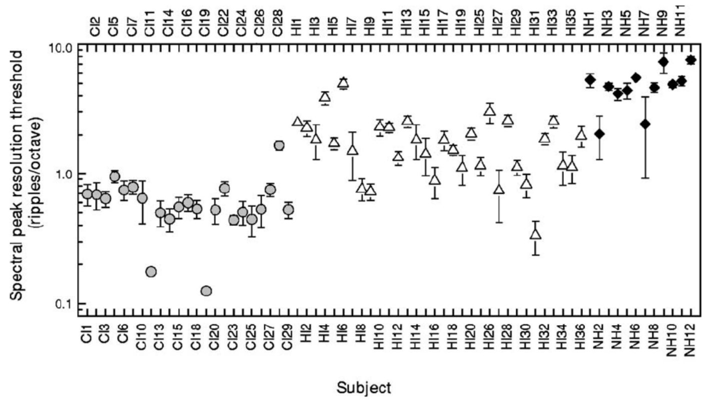
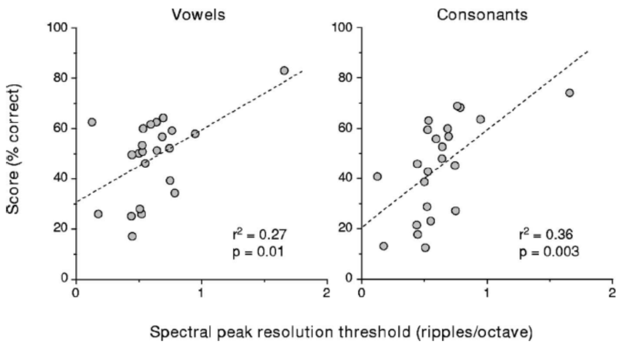


FIG. 2. Thresholds for spectral peak resolution for NH, HI, and CI subjects. Error bars represent ± one standard deviation.

Henry *et al.* 2005 J Acoust Soc Am

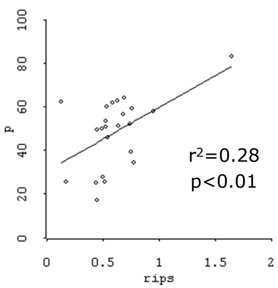
Relationships to performance with speech in quiet

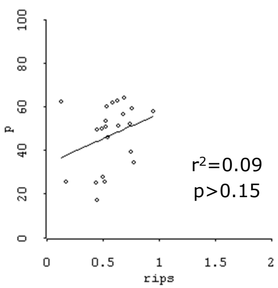
12 hVd by 20 talkers 16 VCVs by 4 talkers



Henry *et al.* 2005 J Acoust Soc Am

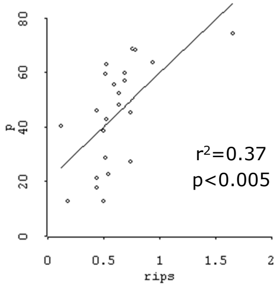
Statistical interlude: The effect of outliers

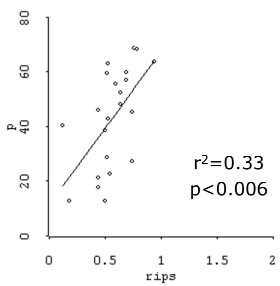




vowels

Statistical interlude: The effect of outliers





consonants

Relationships to performance with speech in noise

SRT determined for selection of one of 12 spondees

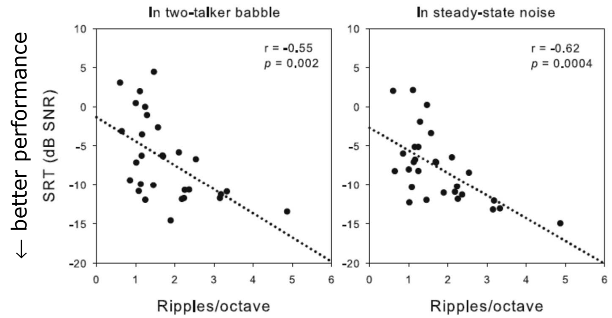


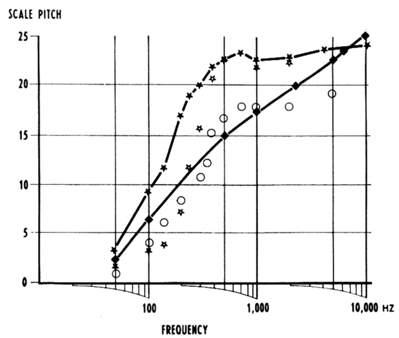
FIG. 6. Spectral-ripple discrimination is correlated with speech perception in noise. The figure shows the relationship between the spectral-ripple thresholds and SRTs in two-talker babble (*left panel*) and steady-state noise (*right panel*) using data from the first six repetitions. Linear regressions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

Won *et al.* 2005 JARO

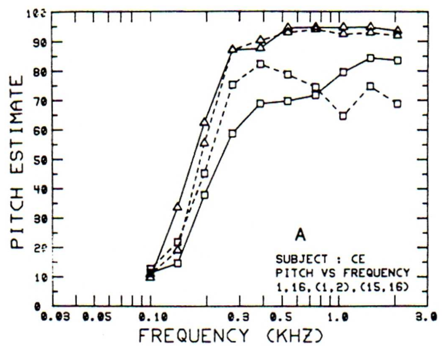
Why is speech melody (*voice pitch)* important to hear?

- 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



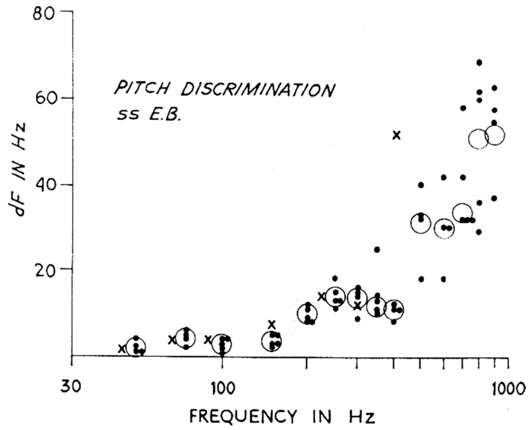
Merzenich *et al.* 1973



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

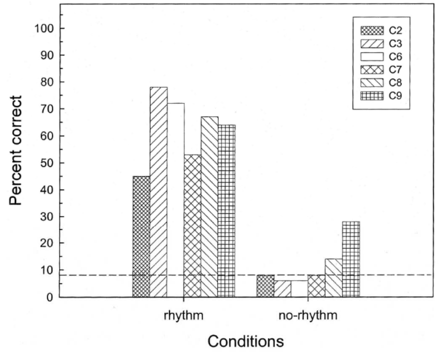
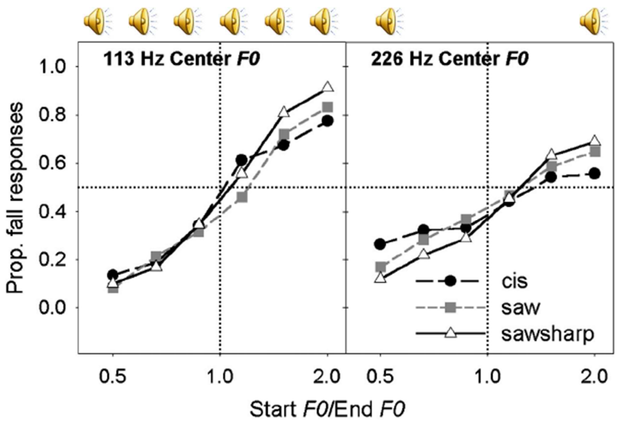


Figure 4. Melody identification scores from individual cochlear implant listeners with the original melodies. The horizontal dashed line indicates the mean chance performance. The vertical bars represent different subjects in each condition.

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

Kong *et al.* (2004)

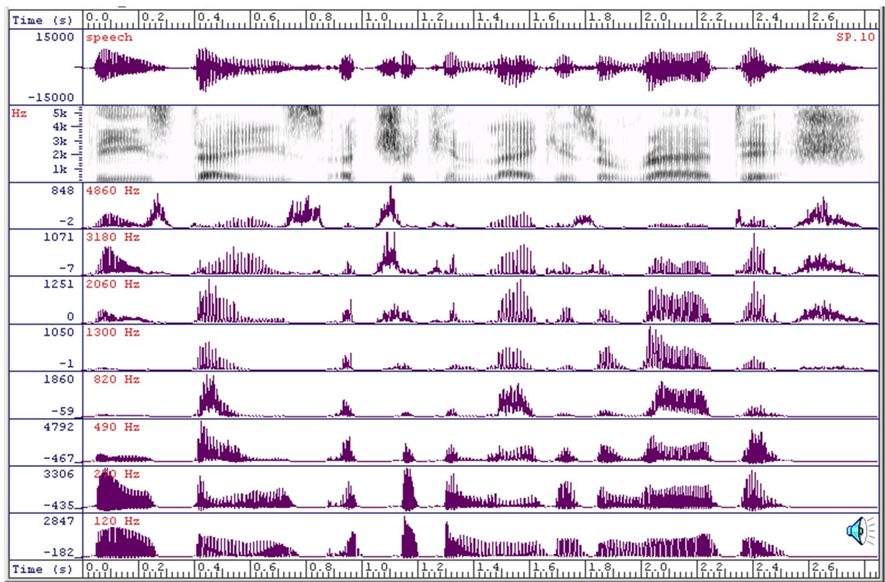
CI users classifying rise/fall contours on diphthongs



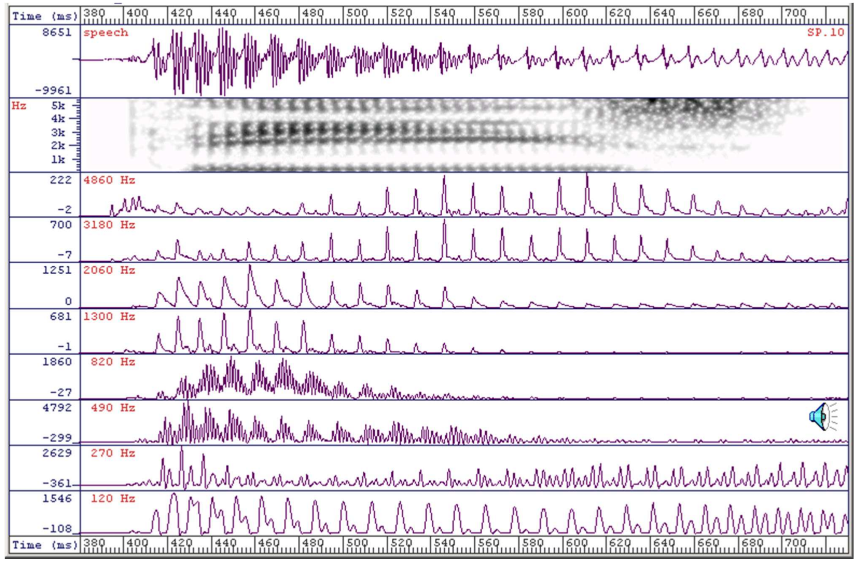
Green *et al.* 2004 J Acoust Soc Amer

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



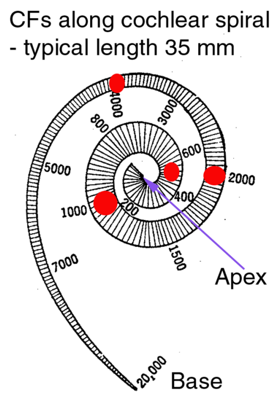
The representation of melody can be messy!



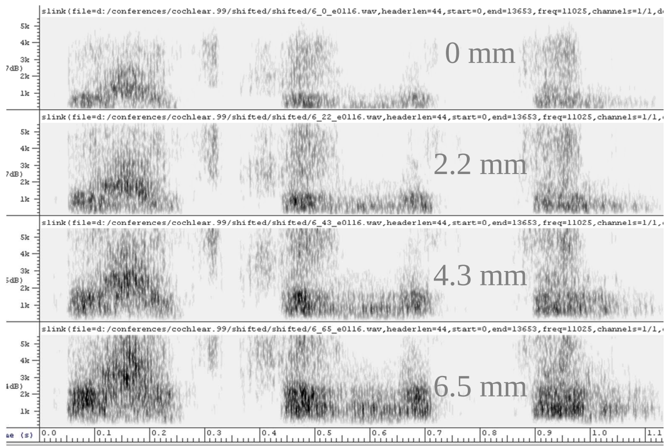
Perception of fundamental pitch in complex waves is very poor

- 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

What happens when an electrode is incompletely inserted?

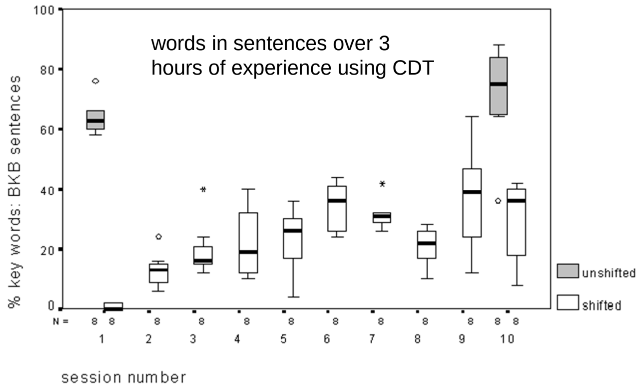


Simulations of incomplete insertions



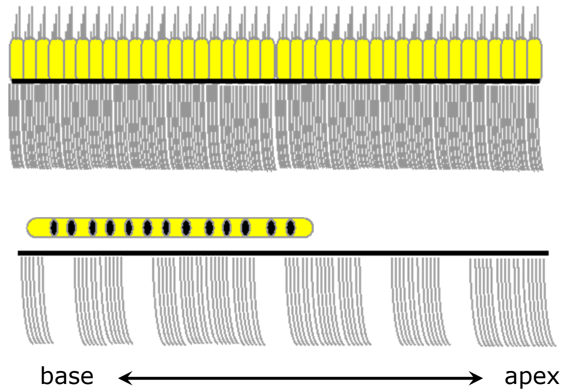
Can the deleterious effects of spectral shifting be overcome over time?

Pre-training ❖ Post-training ❖



normal listeners in simulations: Rosen *et al.* 1999 J Acoust Soc Am

Hair cell substitution?



from Lynne Werner: http://depts.washington.edu/sphsc461/CI_notes.htm

Why is a CI not as good as normal hearing?

- It's a *damaged* auditory system, presumably with accompanying neural degeneration (e.g. dead regions)
- Electrodes may not extend fully along the length of the basilar membrane (BM), so mis-matched tuning and restricted access to apical regions (where nerve survival is typically greatest)
- 3000 IHCs vs. a couple of dozen electrodes, hence poorer frequency selectivity
- Current spreads across BM, hence poorer frequency selectivity
- Less independence of firing across nerve fibres, appears to affect temporal coding
- Small dynamic ranges but intensity jnd's not correspondingly smaller, hence fewer discriminable steps in loudness
- But good temporal and intensity resolution