Role of F0 differences in source segregation

Andrew J. Oxenham

Research Laboratory of Electronics, MIT

and

Harvard-MIT Speech and Hearing Bioscience and Technology Program

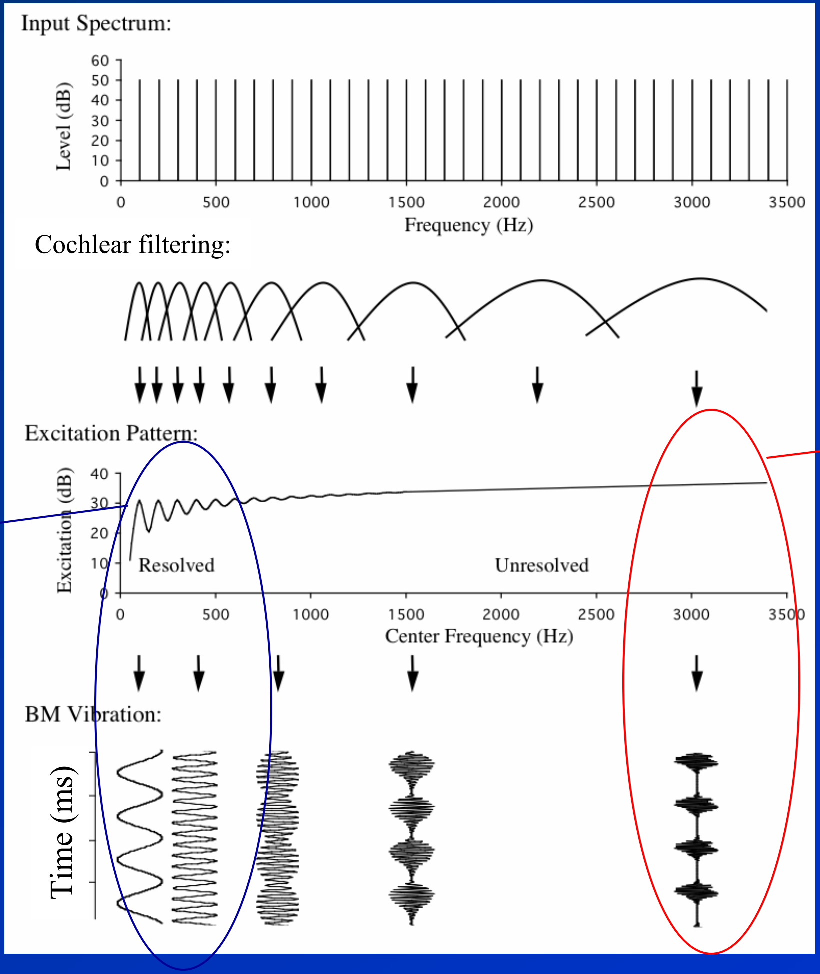
Rationale

- Many aspects of segregation involve monaural cues.
- F0 is one of the most obvious and most studied.

How robust are these cues, and which pitch cues are most useful?

Harmonic complex tones

Many sounds in our world are harmonic complex tones, consisting of many sinusoids all at multiples of the *fundamental frequency* (F0).



Resolved harmonics: Temporal fine structure

Unresolved harmonics: Temporal envelope

(Plack & Oxenham, 2004)

High (unresolved) harmonics produce poor musical pitch

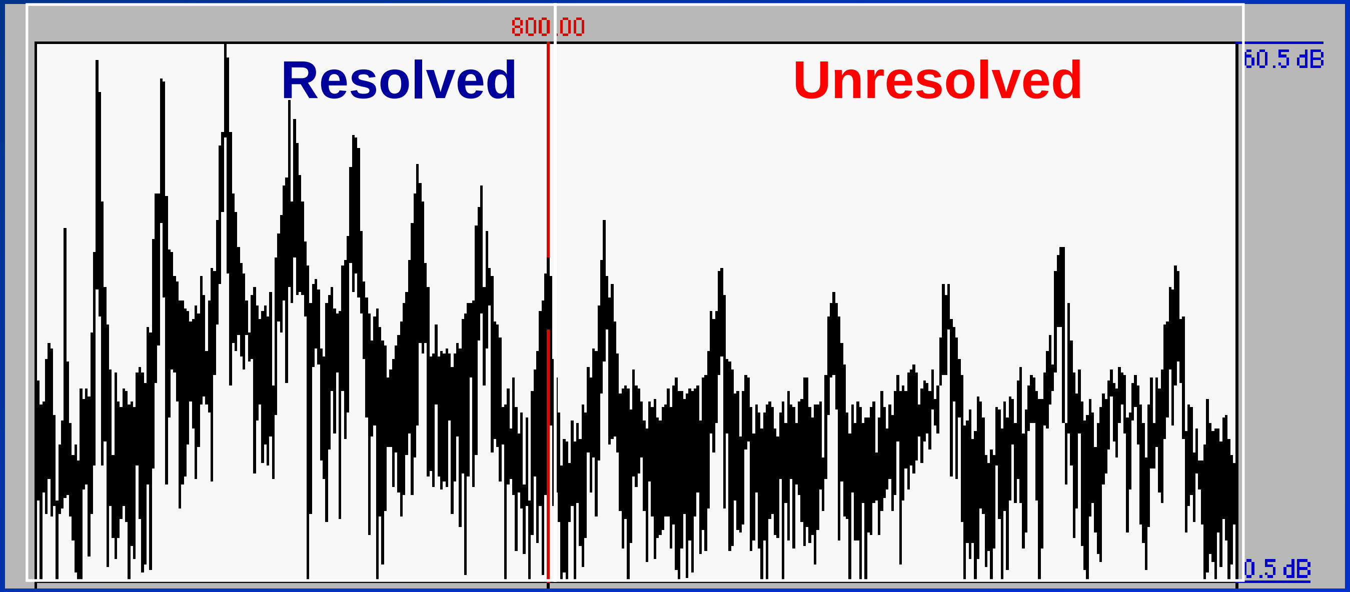
Unresolved	Highpass filtered above 8th harmonic
Resolved	Lowpass filtered below 8th harmonic
Resolved & Unresolved	No filtering

(Thanks to Bertrand Delgutte)

Low (resolved) harmonics dominate pitch perception

100-100 100-106 100-112 100-133 100-178

F0 below 800 Hz F0 above 800 Hz



Resynthesized sentences with low- and high-spectral regions on different F0s (Demo by C.J. Darwin)

What we know about pitch coding

Low harmonics

- Spectrally resolved
- Temporal fine structure
- Strong pitch percept

High harmonics

- Spectrally unresolved
- Temporal envelope
- Weak pitch percept

Good pitch perception requires low, spectrally resolved harmonics, represented by their temporal fine structure.

Why may fine structure be important for speech?

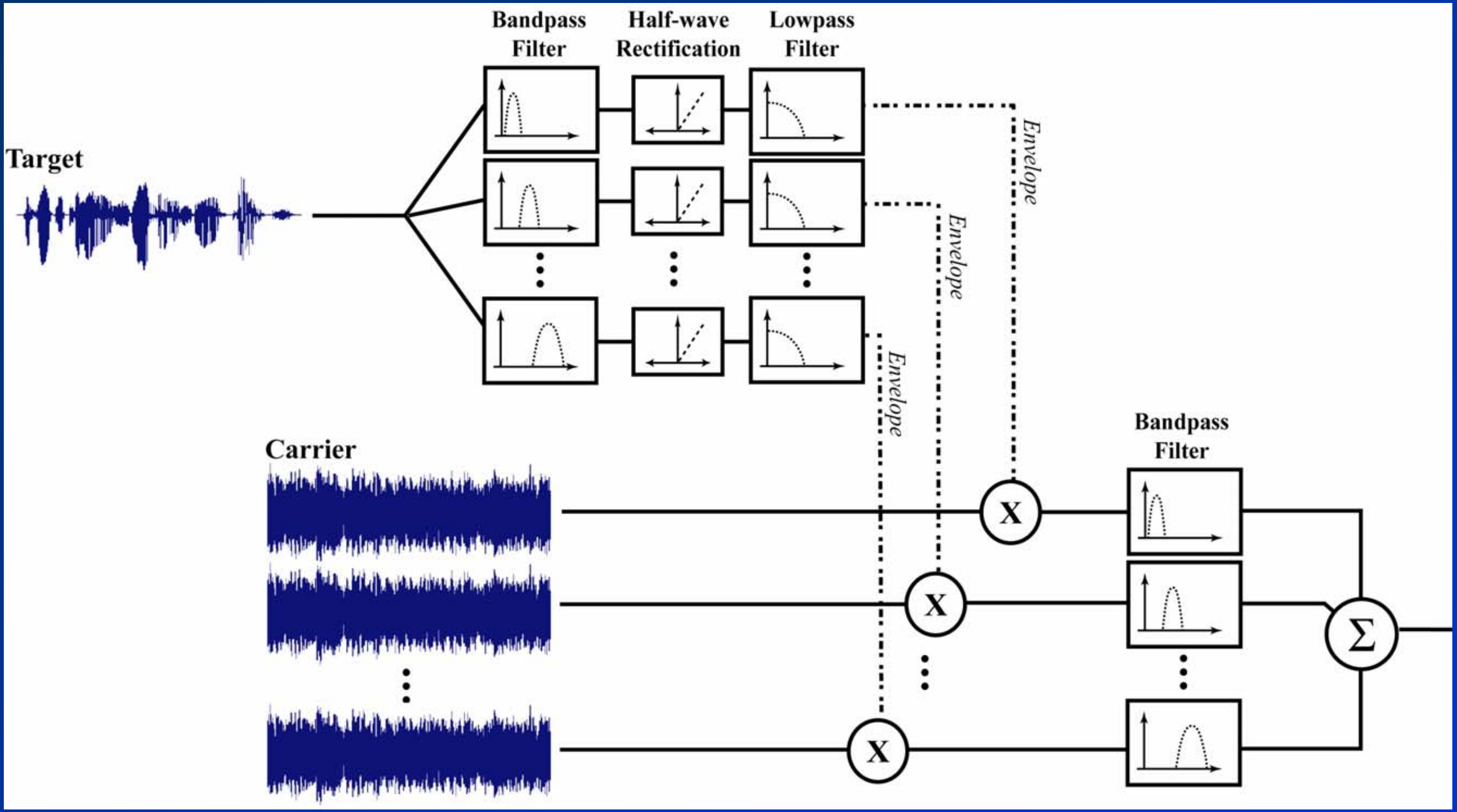
Potential reasons:

1. Good pitch perception needed for prosody (and lexical) information.

2. More robust against reverberation effects.

3. Important for source segregation

Exploring the role of fine structure



Simulates aspects of cochlear implant processing by limiting frequency resolution and replacing original fine structure with noise. (e.g., Shannon et al., 1995)

Using F0 differences

- Small F0 differences (< 1 ST) can be detected, even with small numbers of channels in CI simulations.
- Can these detectable differences in F0 be used for (simultaneous) source segregation?
- Can a reintroduction of some very low-frequency fine-structure information help?

Double-vowel experiments

The ability to hear out two simultaneous vowels improves with F0 difference.

(e.g. Assmann & Summerfield, 1994)

V1+V2 V1+V2 with F0 difference V1 V2

- Synthesized stimuli; artificial presentation

But

- All other cues (onset differences, vocal tract size, dynamic cues) controlled
- Only F0 differences remain

Effect of F0 differences in vowel identification

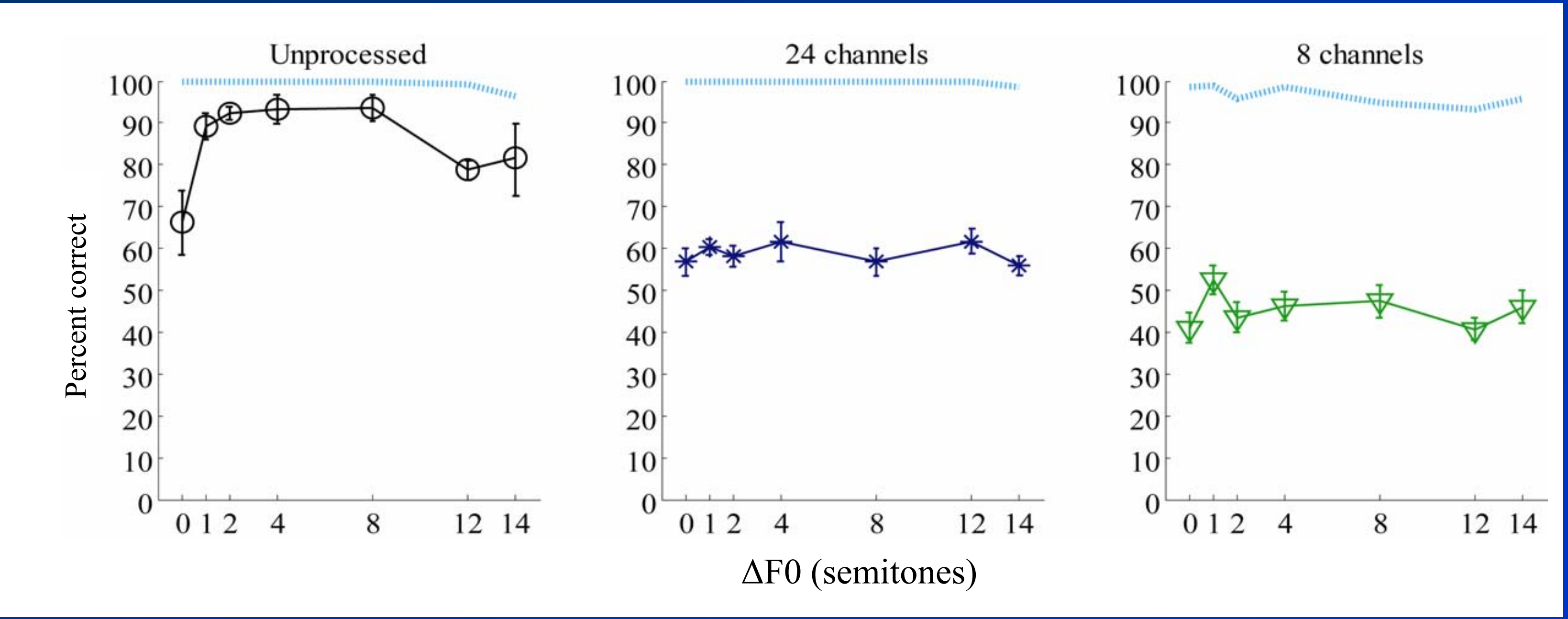
- Stimuli: 5 American-English vowels, presented alone or in pairs.
- Subjects identify as many vowels as possible.
- Processing:
  - Unprocessed
  - CI simulations, with 24 or 8 channels.
- ‘Correct’ only if both vowels correctly identified.

Effects of adding low-frequency information

- Double-vowel experiment
  - 8-channel Noise-excited vocoder (NEV) +
  - Lowpass-filtered (LPF) acoustic information (300 Hz or 600 Hz cutoff)

- Conditions:
  - Just NEV
  - Just LPF
  - NEV + LPF

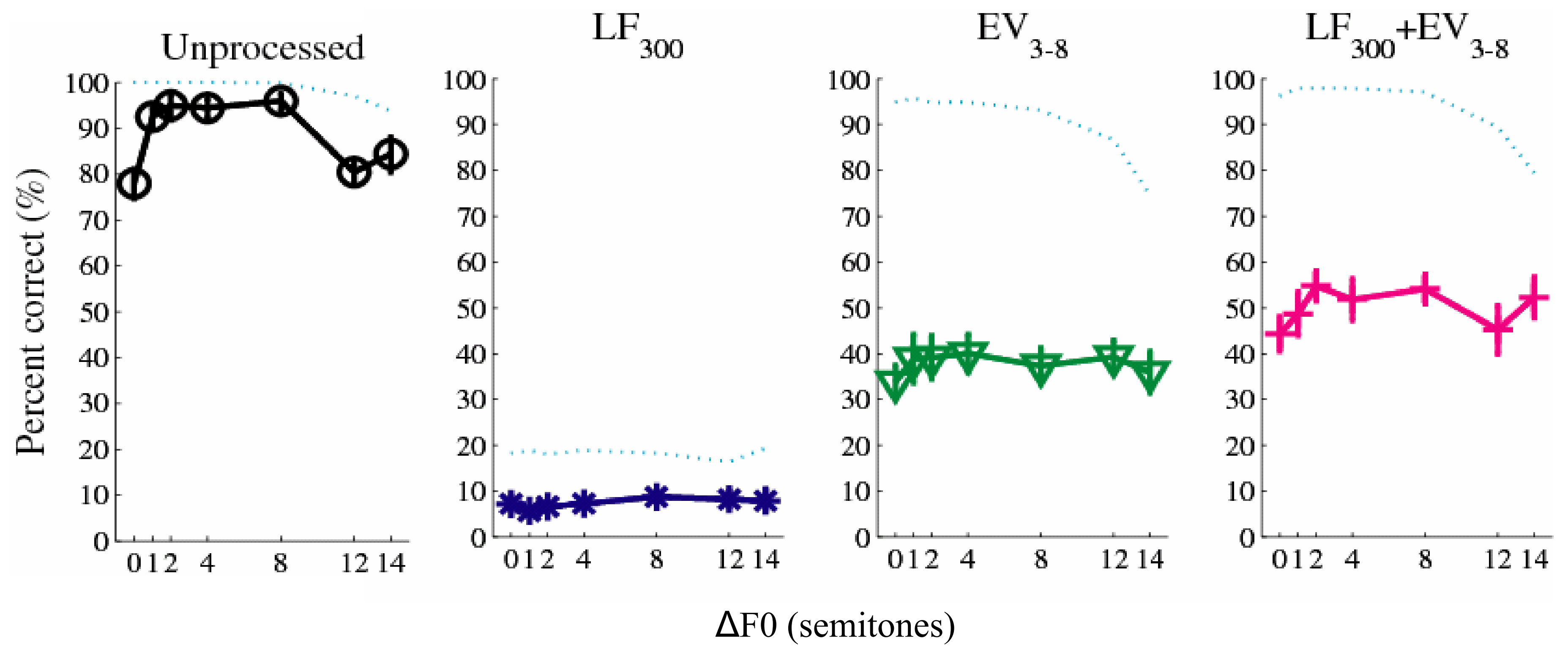
Double-vowel results



- Unprocessed shows benefit of F0 differences, up to 2 semitones.
- Processed conditions show *no benefit* of F0 differences, even with 24 channels.

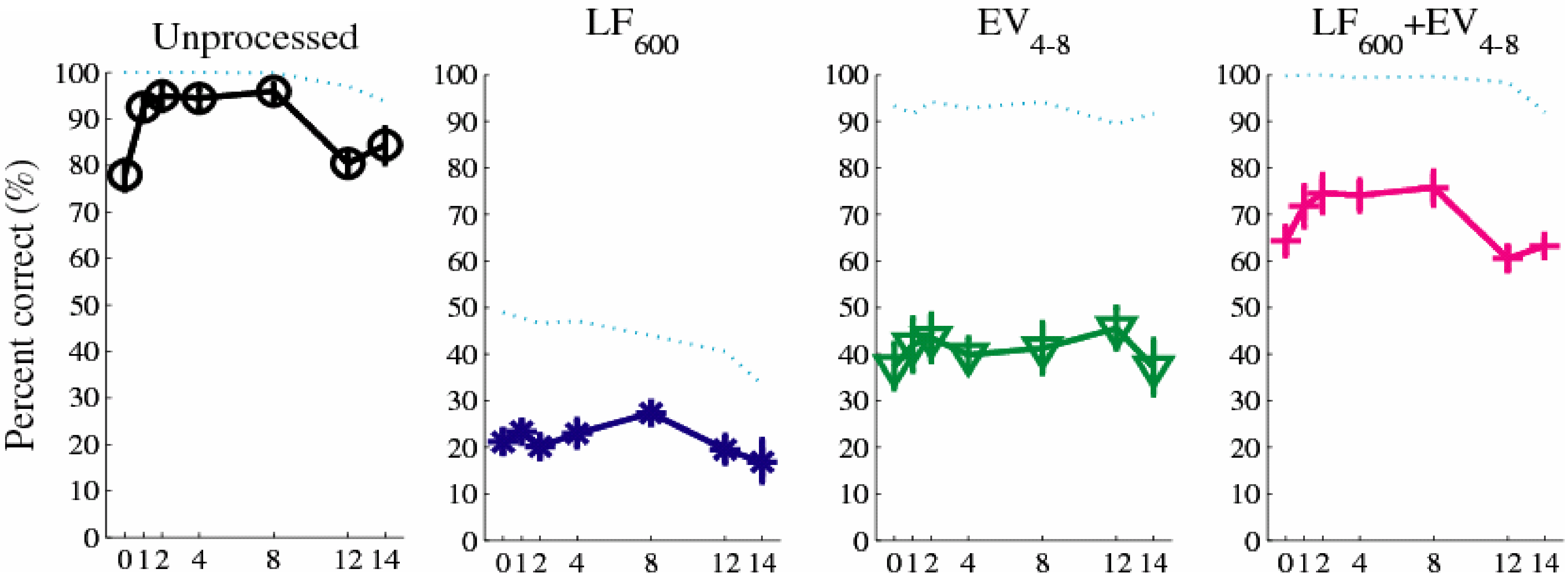
(Qin & Oxenham, 2004)

Double-vowel results: 300 Hz LPF



(Qin & Oxenham, 2004)

Double-vowel results: 600 Hz LPF



(Qin & Oxenham, 2004)

Double-vowel results

- For CI simulations, sequential F0 differences can be detected, but simultaneous F0 differences cannot be exploited to assist in vowel segregation.
- Consistent with results of Carlyon (1996), who found that simultaneous tone complexes in the same spectral region were not heard as two sounds, if they only consisted of high numbered (>10) harmonics.
- The results extend this by showing similar effects even without perfect spectral overlap.

Double-vowel results

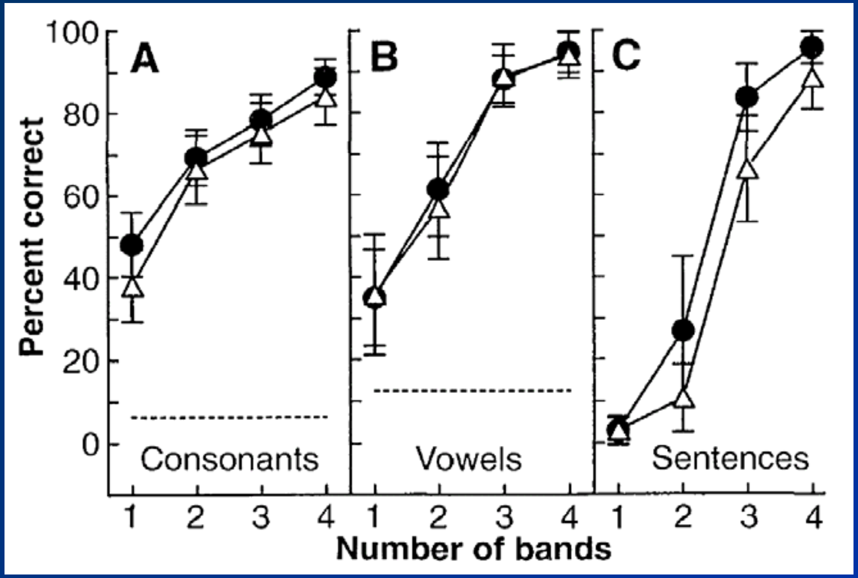
- Reintroducing fine structure below 300 Hz already improves performance somewhat, and leads to benefits of F0 difference (at least in these data).
- Increasing the cut-off frequency to 600 Hz improves performance (dominance region of pitch, or simply more F1 information?)

Residual low-frequency hearing may provide an important supplement to cochlear-implant perception.

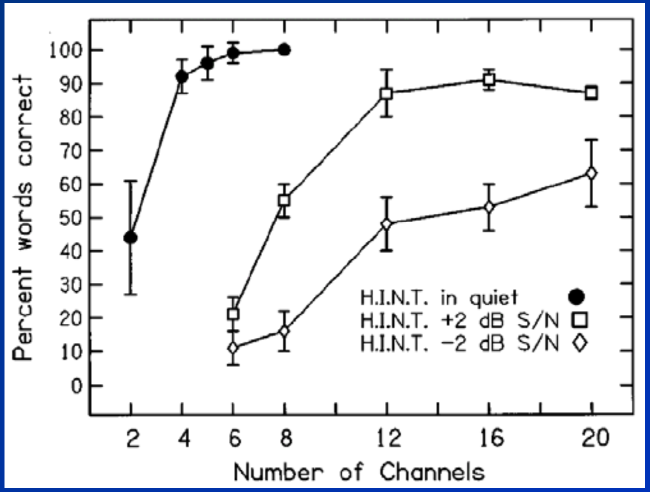
What about “real” speech?

- Normal-hearing listeners show a large release from masking in spectro-temporally complex maskers, compared to steady-state noise.
- Impaired listeners do not.
- Loss of frequency selectivity and/or deterioration in F0 coding?
- Noise-vocoder simulations can (to some extent) distinguish.

Previous CI simulation studies



Shannon et al. (Science, 1995)

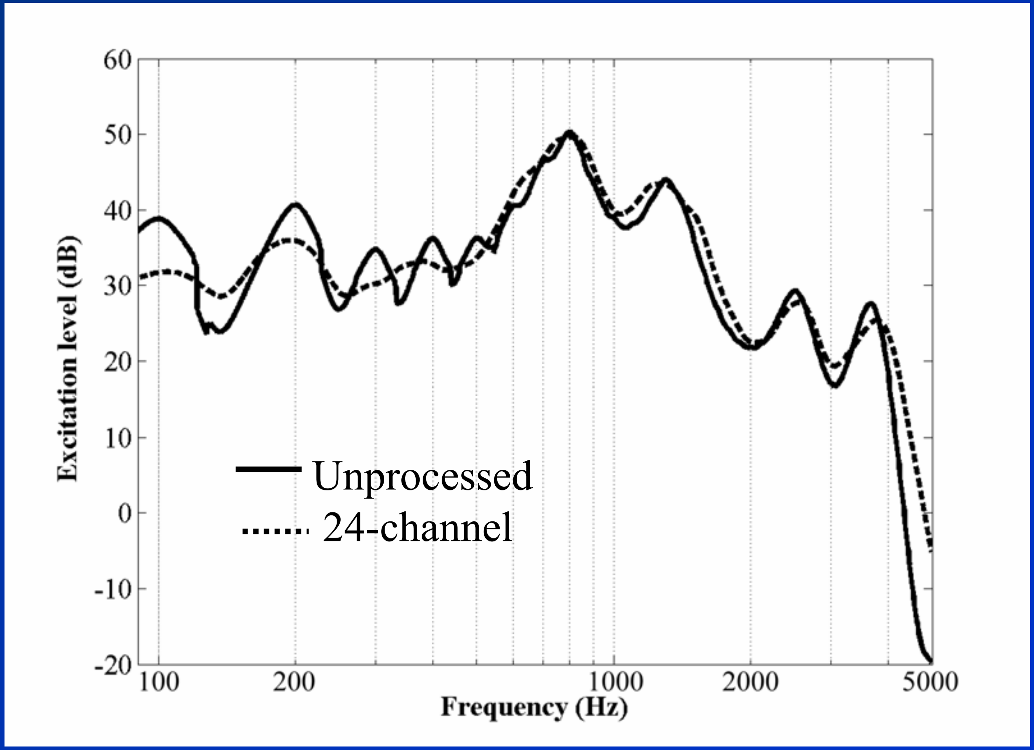


Dorman et al. (JASA, 1998)

- Reasonable speech perception in quiet requires only 4 channels.
- Speech in noise also possible, but with more channels.

Channel numbers

- 4-6 channels: Maximum number of effective channels currently available in CIs.
- 24 channels: similar formant resolution as found in normal hearing.



Noise-excited vocoder examples

- 4 channels in steady noise (0 dB SNR)
- 8 channels in steady noise
- 24 channels in steady noise
- Unprocessed in steady noise

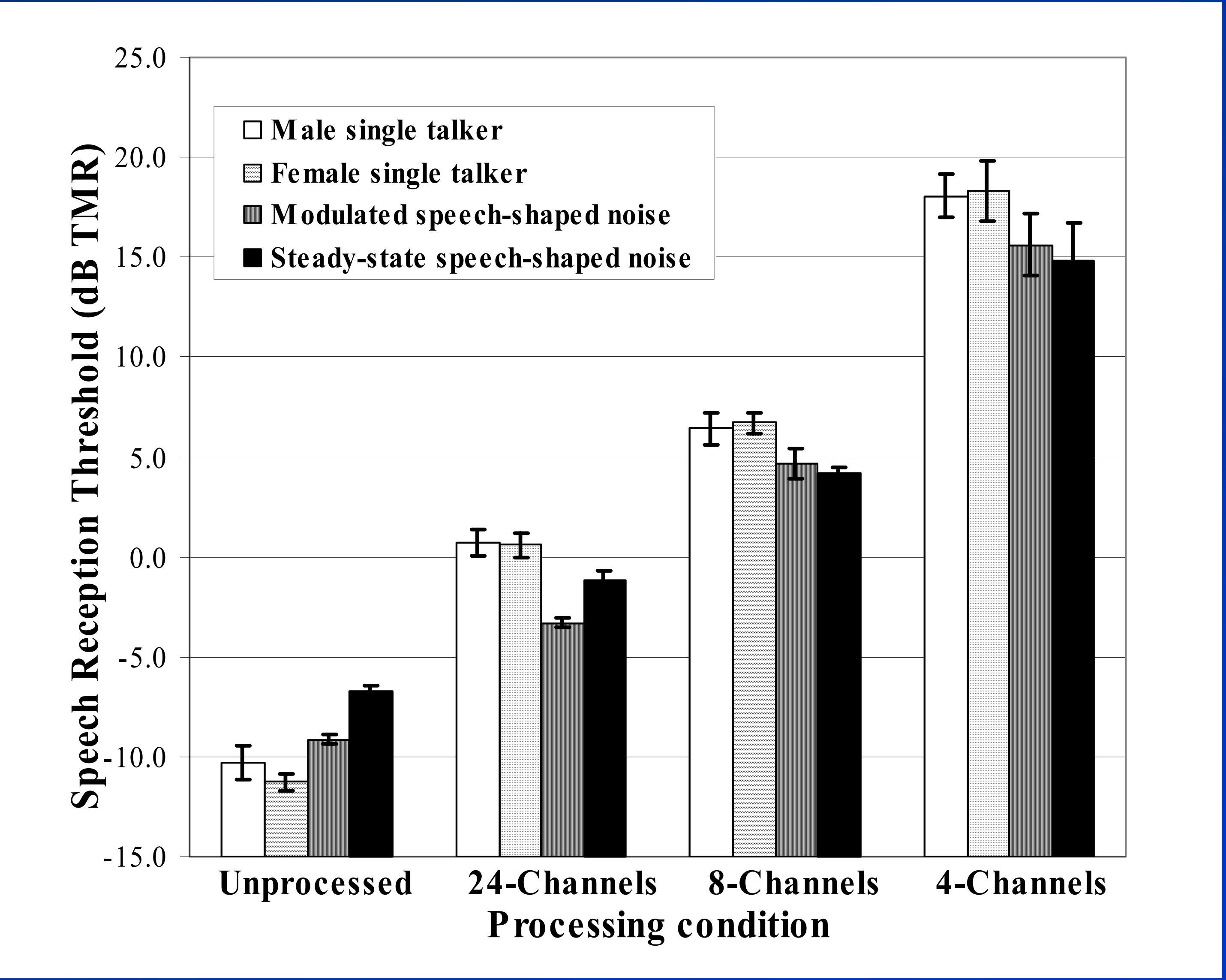
Implant simulations

HINT Sentence recognition

- Backgrounds:
  - – Speech-shaped steady noise
  - – Modulated speech-shaped noise
  - – Single-talker interference (Male and Female)
- Simulated Cochlear Implant Processing:
  - – Noise-excited vocoder (NEV)
  - – Unprocessed, 24, 8, and 4 channels

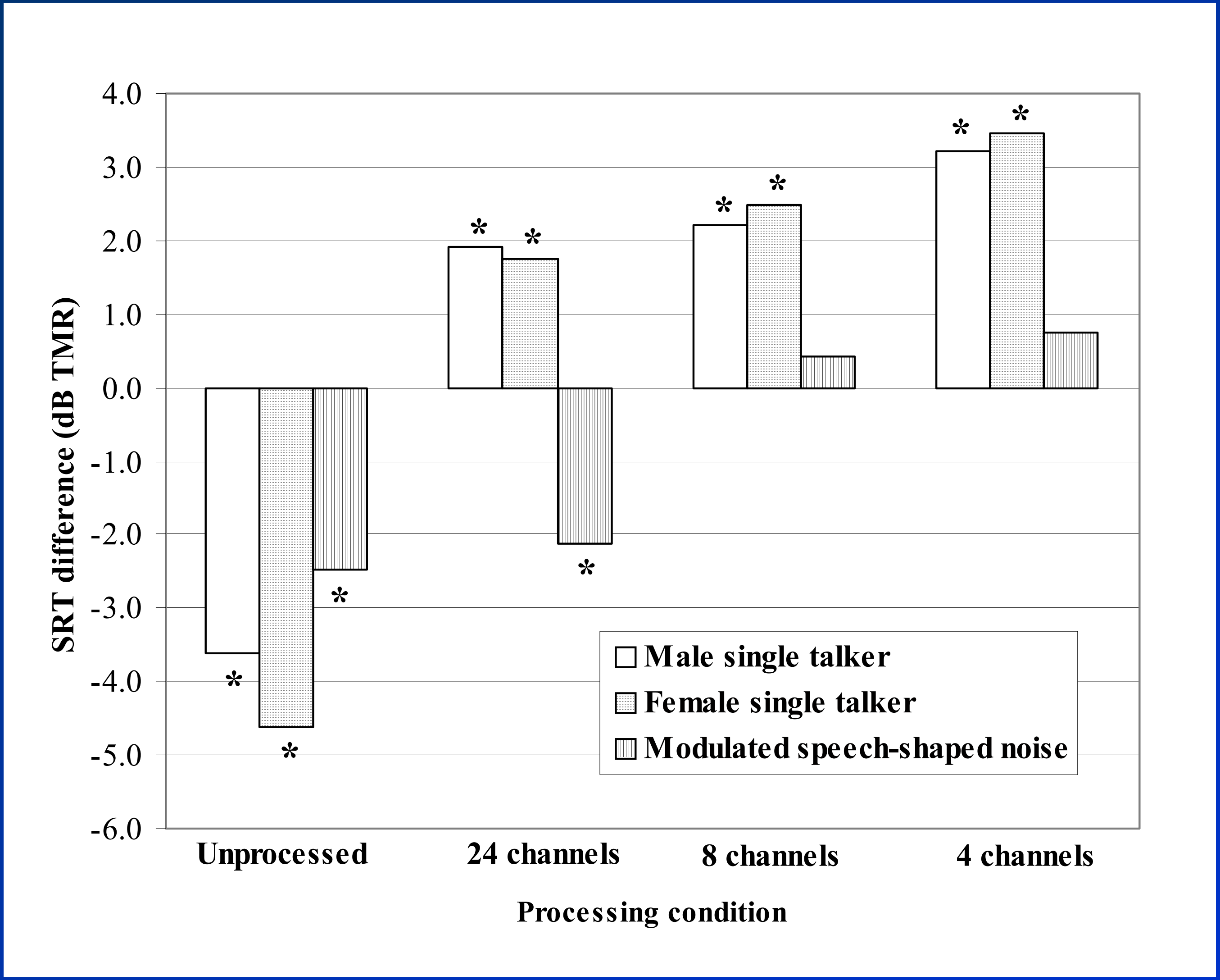
Qin & Oxenham (JASA, 2003)

Simulation results. I



(Qin & Oxenham, 2003)

Simulation results. II

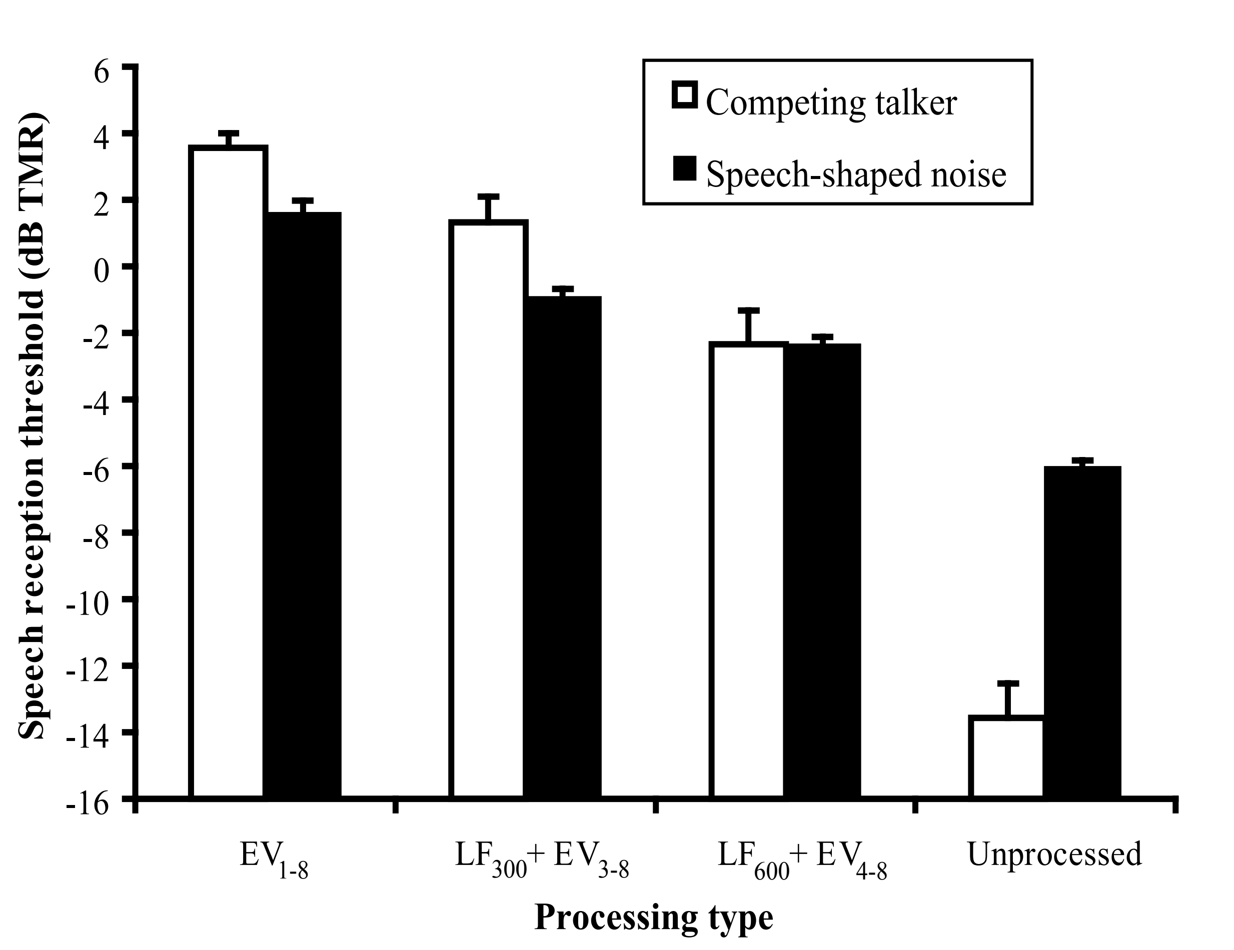


(Qin & Oxenham, 2003)

Effects of implant simulations

- Single-talker went from least effective masker (unprocessed) to most effective masker (processed), even with 24 channels.
- Based on earlier experiments, this may be due to loss of fine-structure cues and pitch.

Reintroducing low-frequency fine structure



Reintroducing fine structure

- Even information below 300 Hz had a positive effect on speech reception, despite no intelligibility alone.
- Improvement with increase of low-frequency cutoff to 600 Hz probably due to improved pitch and F1 representation.

What's so special about low-frequency harmonics?

- Purely temporal models do not predict an advantage of low-numbered harmonics over high-numbered harmonics.
- Is it peripheral resolvability (Carlyon & Shackleton, 1994) or something that simply covaries with it (Bernstein & Oxenham, 2003; de Cheveigné)?
- Emprical and modeling tests underway using multiple harmonic complexes (Micheyl & Oxenham)

Conclusions

- Temporal fine structure is not necessary for speech understanding in quiet, but may be crucial in more complex environments.
- Hearing-impaired listeners rely more on envelope, and cochlear-implant users rely solely on (weak) envelope pitch. This may account for many difficulties in noise.
- Reintroducing some low-frequency information through aided acoustic stimulation may improve performance of cochlear-implant users.

Acknowledgments

Thanks to:

Michael "Q" Qin

Josh Bernstein

Christophe Micheyl

Work supported by the National Institute on Deafness and other Communication Disorders (Grant R01 DC 05216).