

# *Effects of Cross-Spectral and Cross-Modal Temporal Synchrony on Speech Perception*

Ken W. Grant

Joshua G. W. Bernstein

Walter Reed Army Medical Center, Washington, DC 20307

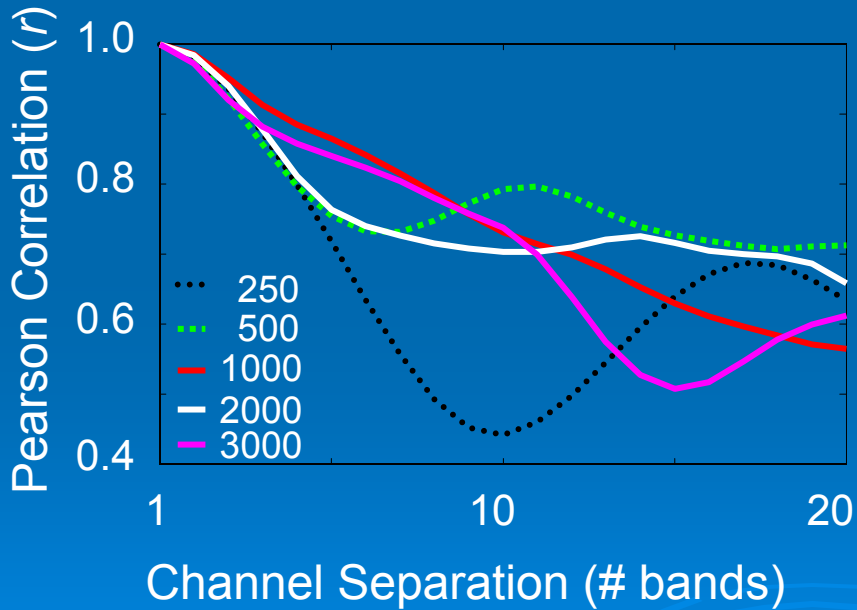
A decorative graphic consisting of several sets of concentric circles, resembling ripples in water, located in the bottom right corner of the slide.

# *Synchronous Characteristics of Dynamic Speech Signals*

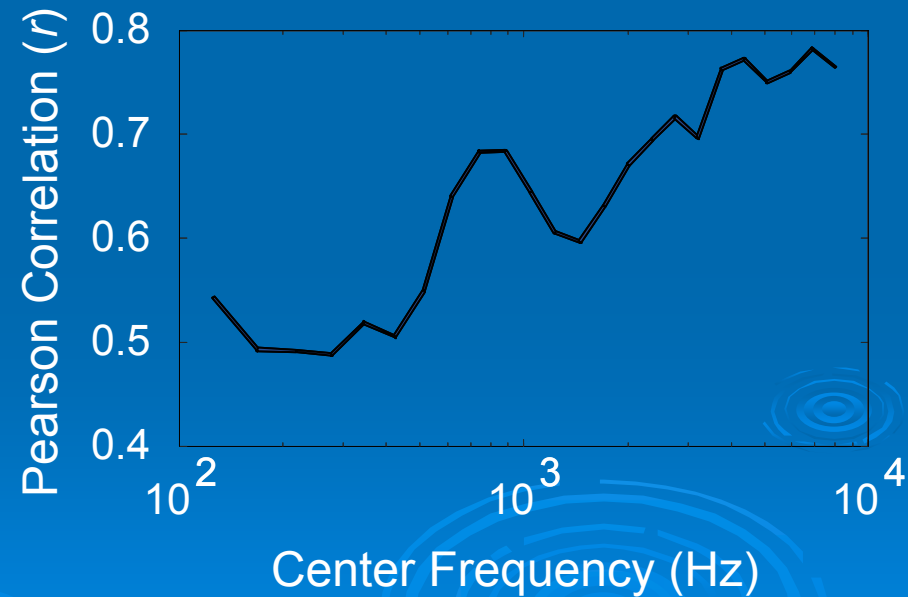
- Inherent synchrony (and asynchrony) across spectral channels and across auditory (A) and visual (V) modalities provide important cues to speech perception
- Synchrony across different sources of speech information can impact speech perception at many levels
  - Speech Recognition
  - Speech Detection

# *Inherent Synchrony for Auditory and Auditory-Visual Speech*

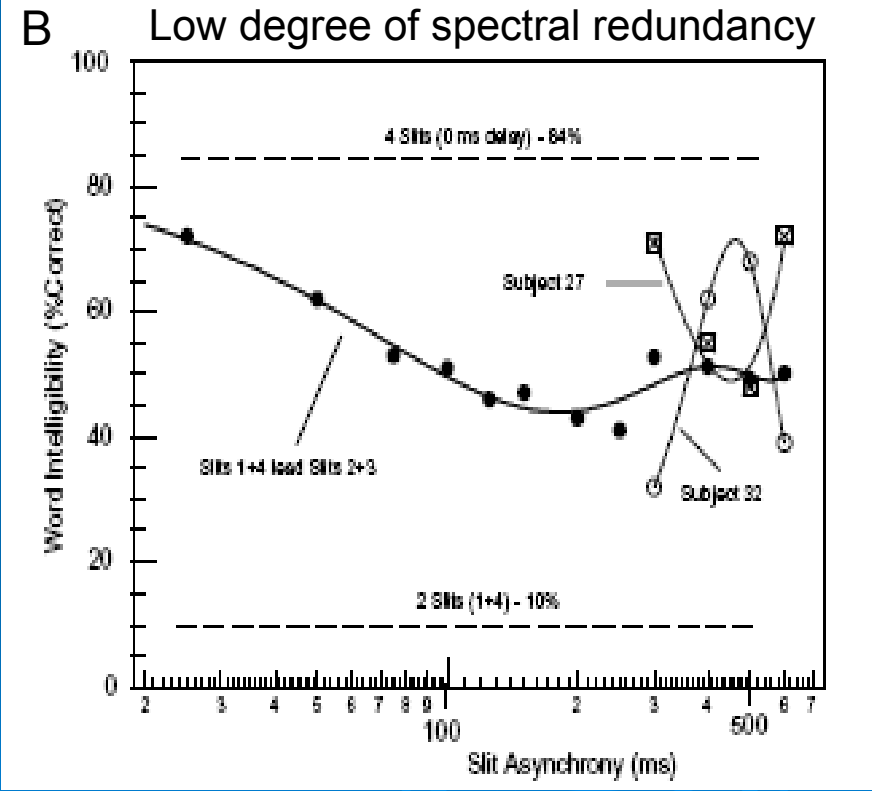
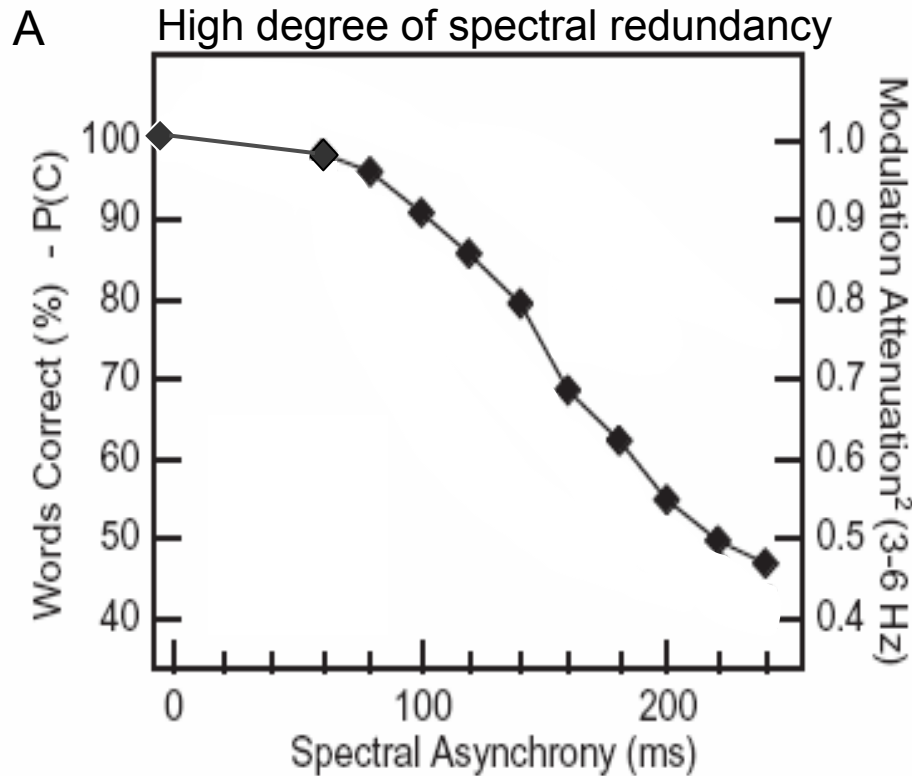
Correlation for speech envelopes as a function of cochlear distance



Correlation for speech kinematics and speech envelope as a function of CF

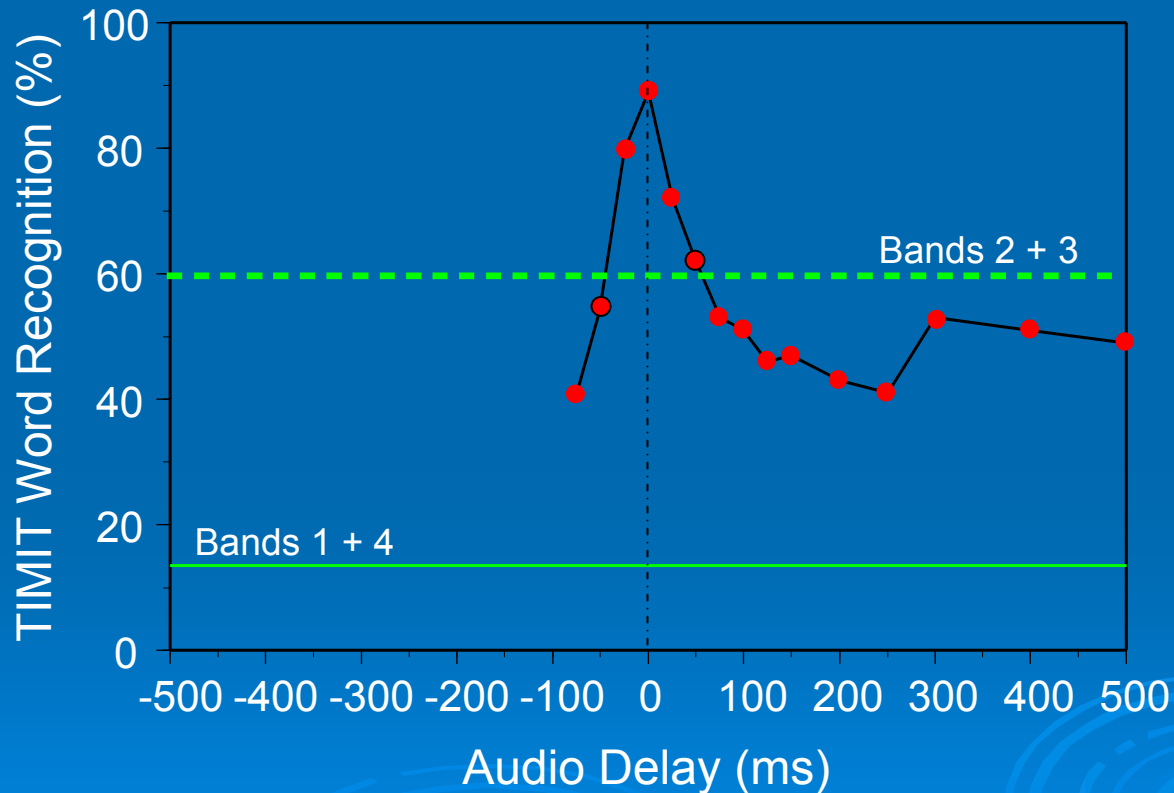


# Effect of Spectral Asynchrony on Auditory Speech Intelligibility



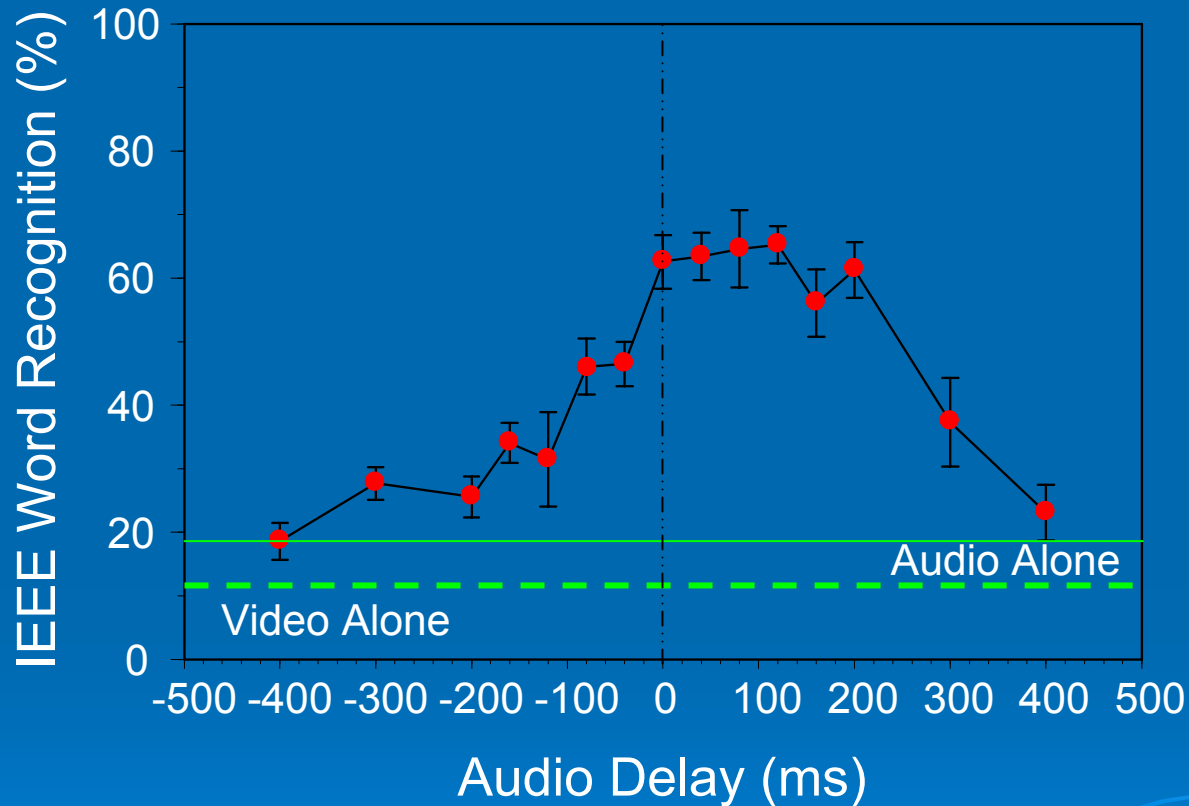
- A. Adapted from Greenberg and Arai, 1998
- B. Adapted from Silipo et al., 1999

# *Effect of Spectral Asynchrony on Auditory Sentence Intelligibility*



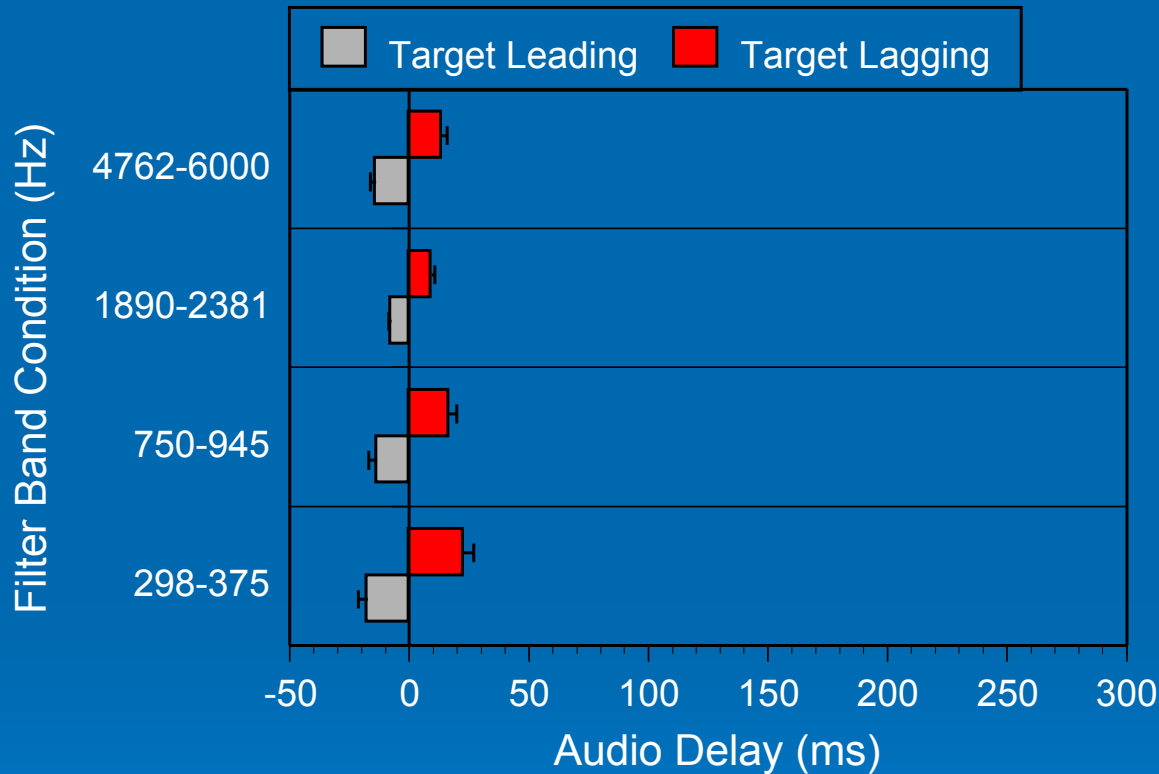
Temporal integration window is symmetric and narrow.

# *Effect of Spectral Asynchrony on Auditory-Visual Sentence Intelligibility*



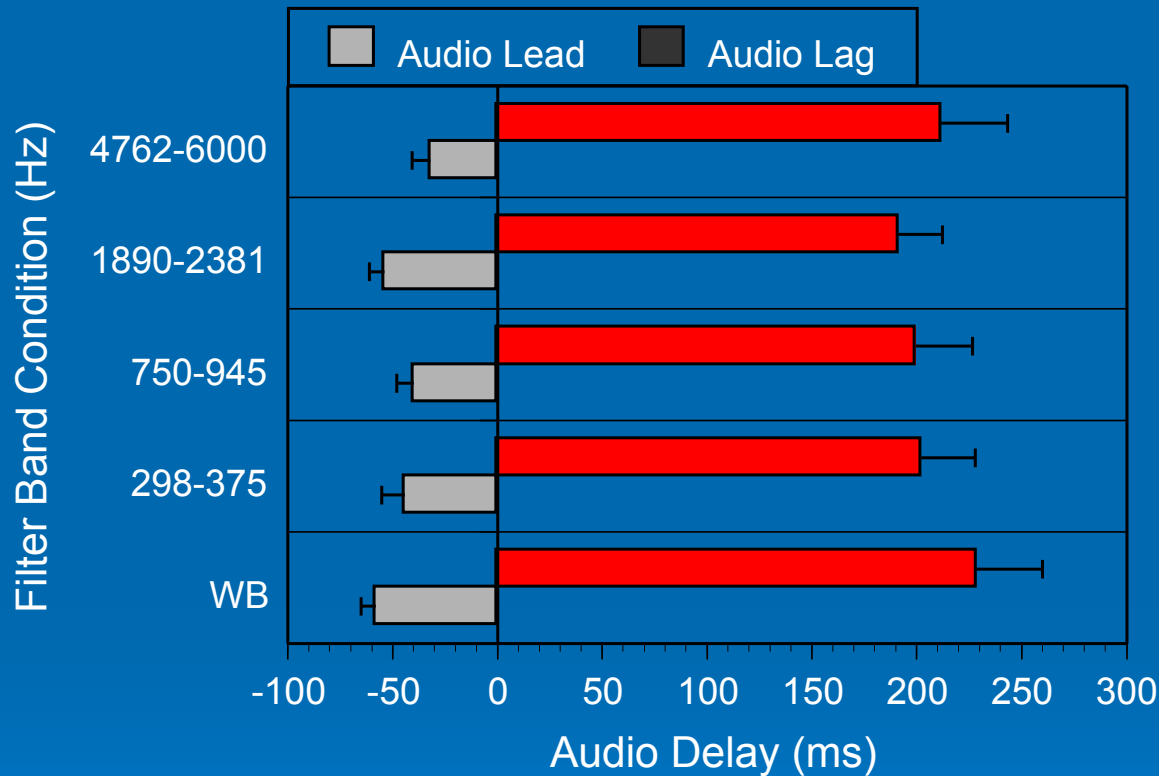
Temporal integration window is asymmetric and broad.

# Detection of Cross-Spectral Asynchrony



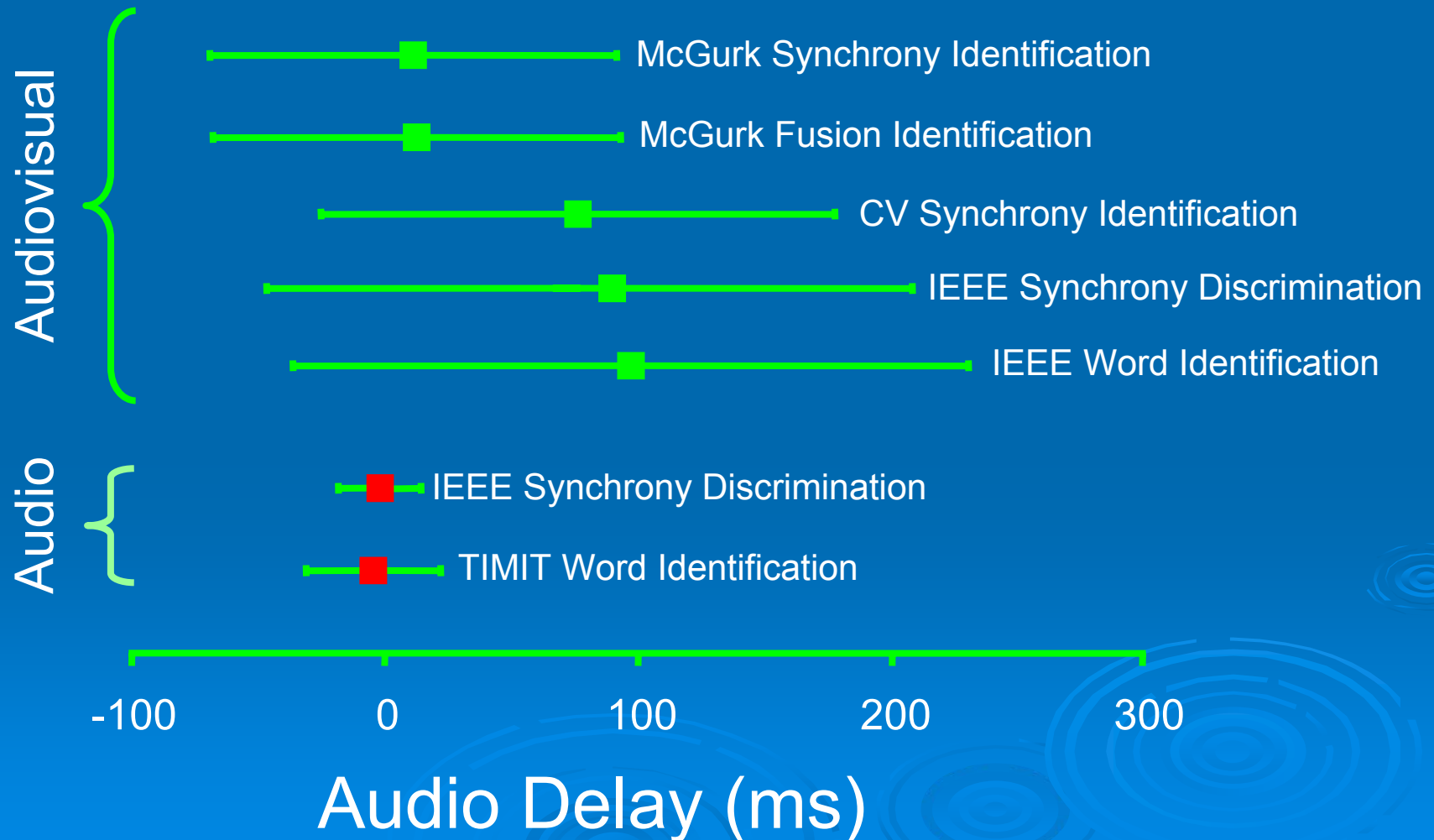
Auditory detection of spectral asynchrony of widely spaced bands is symmetrical and between 10-30 ms (phone level).

# Detection of Cross-Modal Asynchrony



Auditory-Visual detection of a single asynchronous audio channel is asymmetrical and between -60 to 230 ms (syllabic level).

# *Temporal Windows of Processing: Evidence for Multiple Time Frames*



# *Concluding Remarks*

- Perceptual integration windows appear to operate over several (possibly simultaneous?) time frames, both long and short
- Auditory time windows of integration appear to be associated with both syllabic ( $\approx 200$  ms) and phonetic/feature ( $\approx 25$ -50 ms) levels of speech
- Auditory-visual time windows of integration are long ( $\approx 200$  ms) so long as the visual signal precedes the acoustic signal.

# *Thank You!!*

## ➤ Acknowledgments

- Colleagues

- Van Summers      Elena Grassi      Matt Makashay
- Brian Walden      Mary Cord      David Poeppel
- Viginie van Wassenhove      Shihab Shamma
- Steven Greenberg

- Funding

- Oticon Foundation (Graham Naylor)
- GN ReSound (Andrew Dittberner)

# *Extra Slides*



# Summary

- Cochlear channels that are highly correlated (redundant information) provide less speech intelligibility benefit when combined than channels which are weakly correlated (complementary information)
- Desynchronizing spectral bands leads to reduced intelligibility.....but it depends on how the speech signals are constructed
  - 140-200 ms for spectrally dense (redundant) signals (*Arai & Greenberg, 1998*)
  - 25-50 ms for spectrally sparse signals (complementary) signals (*Silipo, Greenberg, & Arai, 1999*)
- For AV speech, intelligibility is stable over very long temporal windows (*Grant et al., 2004*)
  - 200-250 ms (visual speech leads audio speech)
- However....
  - 25-50 ms (audio speech leads visual speech)

# Cross-Spectral Correlation (Speech versus Noise)

