

Spectro-Temporal Interactions in Auditory and Auditory-Visual Speech Processing

Ken W. Grant

Walter Reed Army Medical Center, Washington, D.C.

Steven Greenberg

The Speech Institute, Oakland, CA

<http://www.wramc.amedd.army.mil/departments/aasc/avlab>

grant@tidalwave.net

Acknowledgments

Collaborations:

David Poeppel and Virginie van Wassenhove

Neuroscience and Cognitive Science Program, University of Maryland, College Park, MD

Funding:

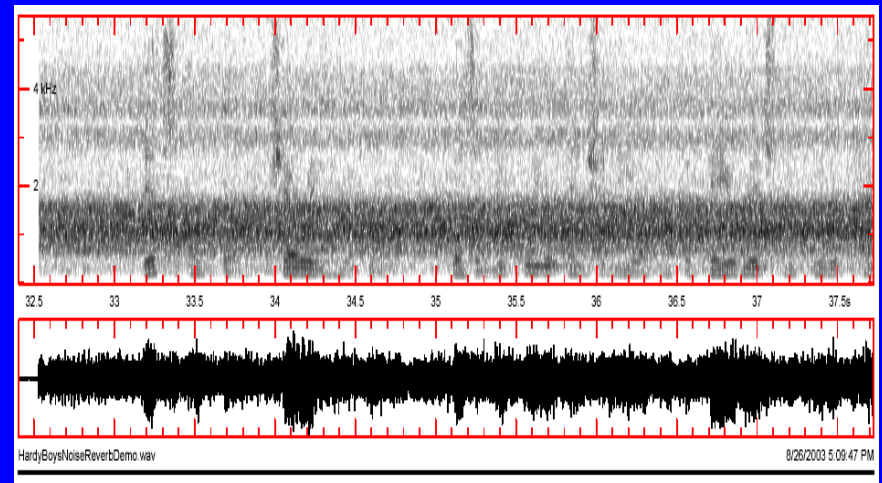
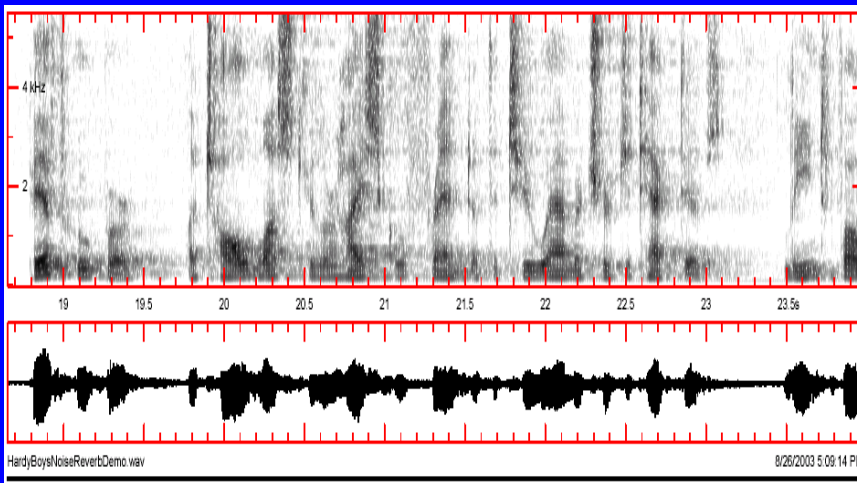
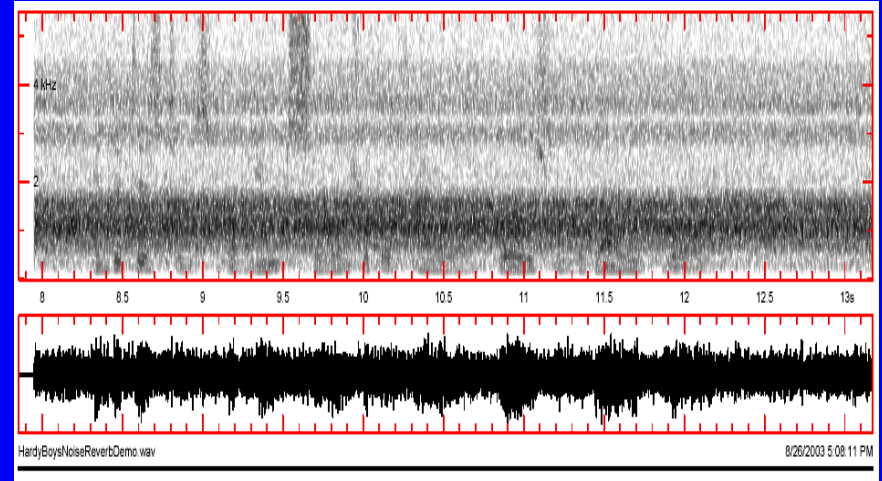
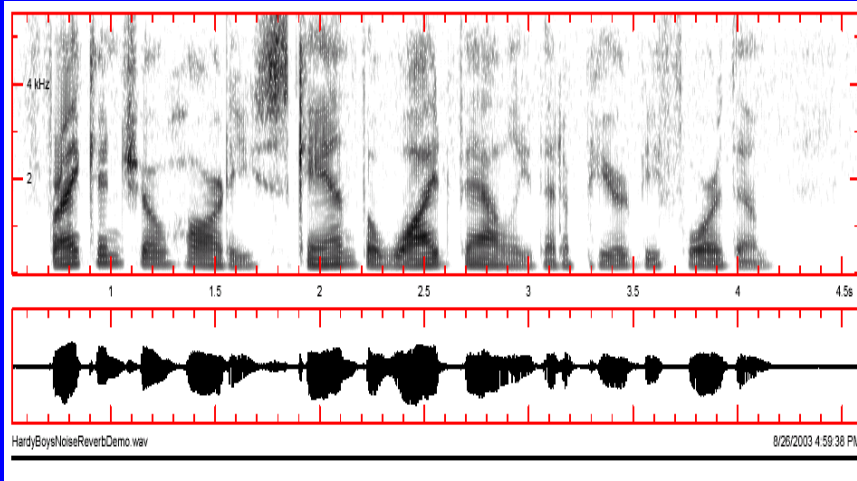
NIH Grant: DC 00792-01A1

NSF Grant (subcontract): SBR 9720398 - Learning and Intelligent Systems Initiative of the National Science

Speech Recognition in Noise and Reverberation

- Primary complaint expressed by hearing-impaired and elderly patients
- Important for machine recognition (ASR)

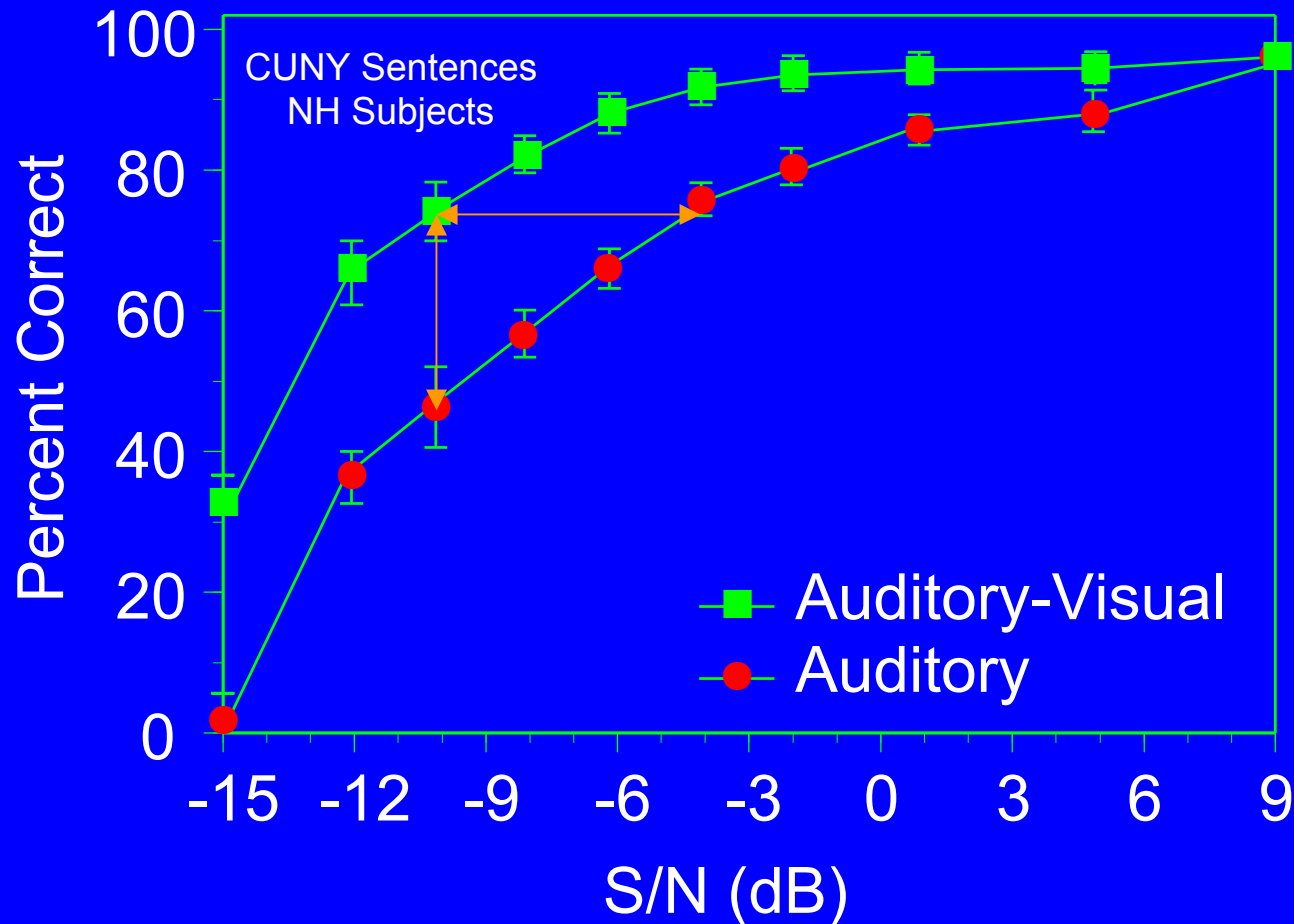
Noisy, Reverberant Speech: Demo



Goal: Improve Speech-to-Noise Ratio

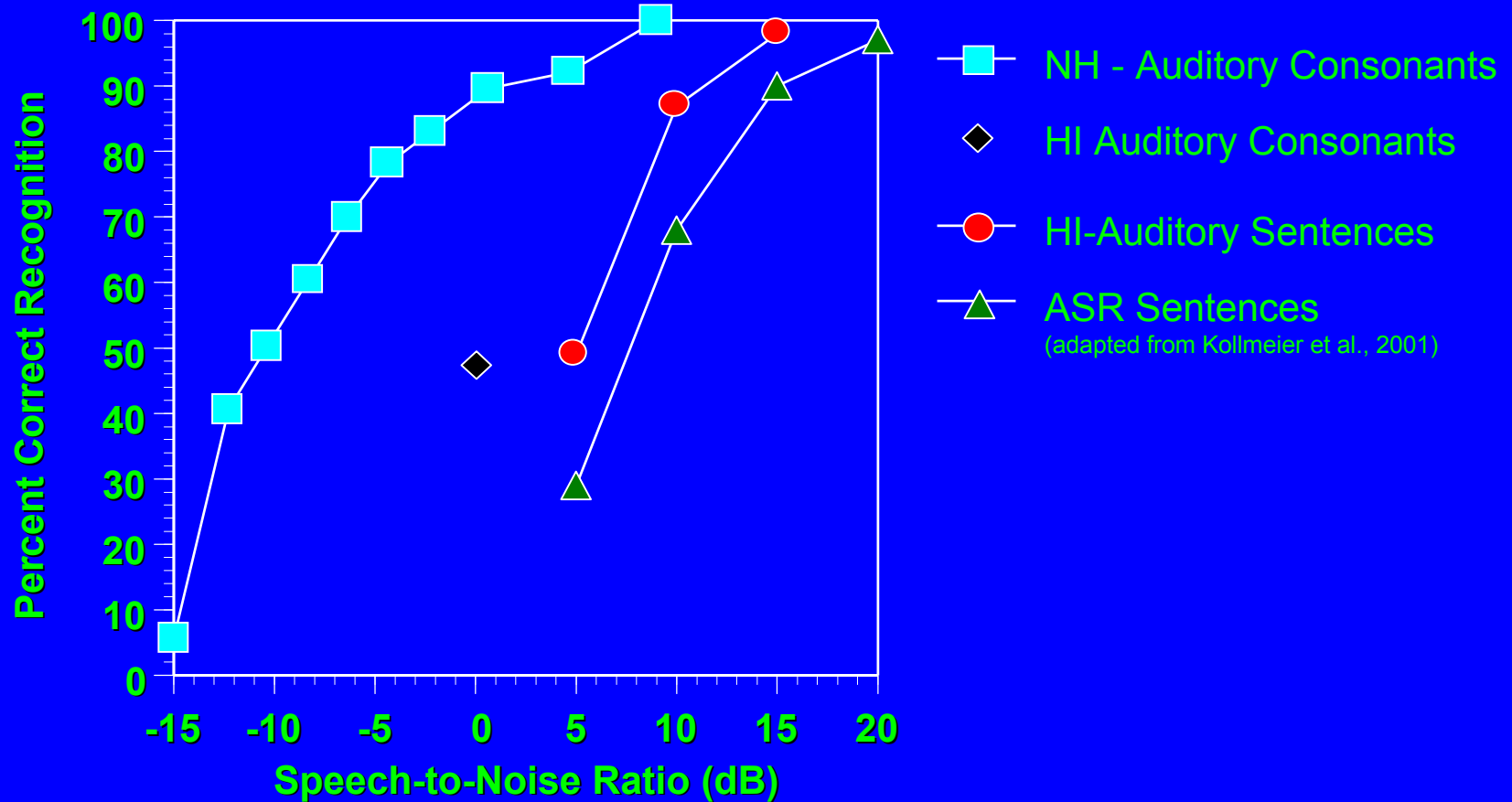
- Signal Processing (e.g., noise reduction algorithms)
- New Technologies (e.g., directional microphones)
- **Speechreading and Auditory-Visual Integration**

Auditory-Visual vs. Audio Speech Recognition

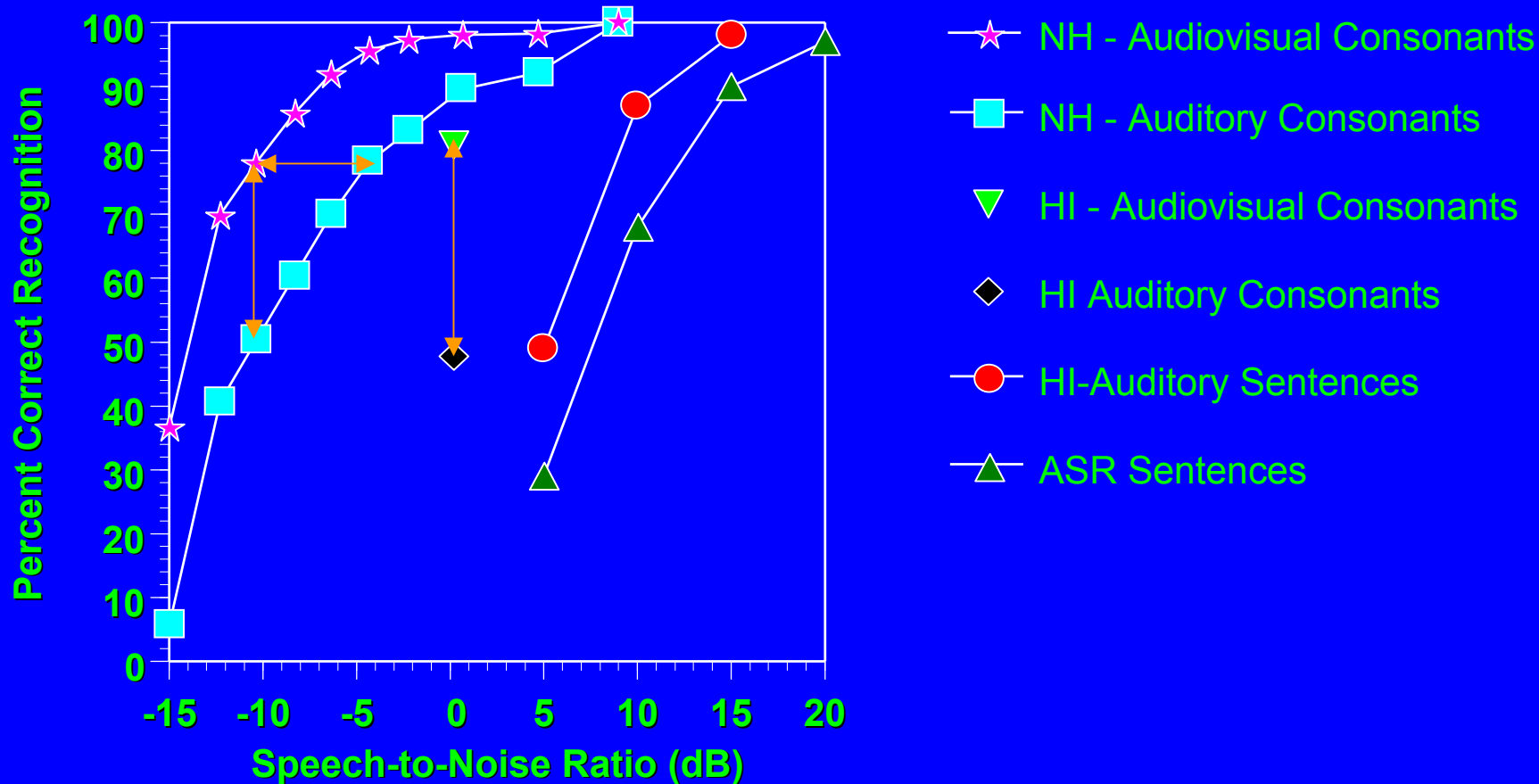


Roughly 6 dB improvement in S/N; roughly 30% improvement in intelligibility for NH subjects

Auditory-Visual vs. Audio Speech Recognition



Auditory-Visual vs. Audio Speech Recognition



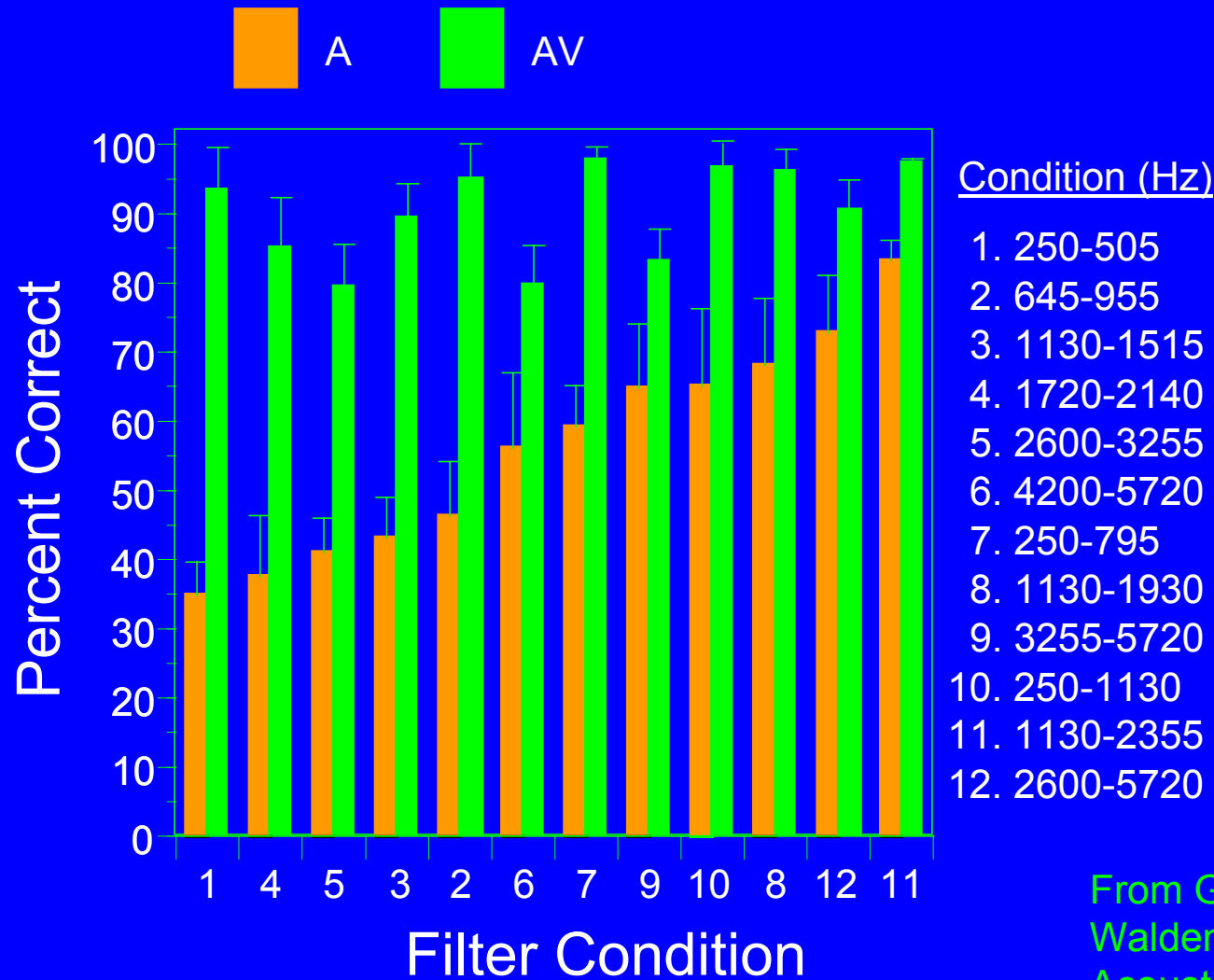
Roughly 6 dB improvement in S/N; roughly 30% improvement in intelligibility for NH subjects.

Spectral Interactions

Audio-visual benefit depends on the spectral locus of the acoustic signal

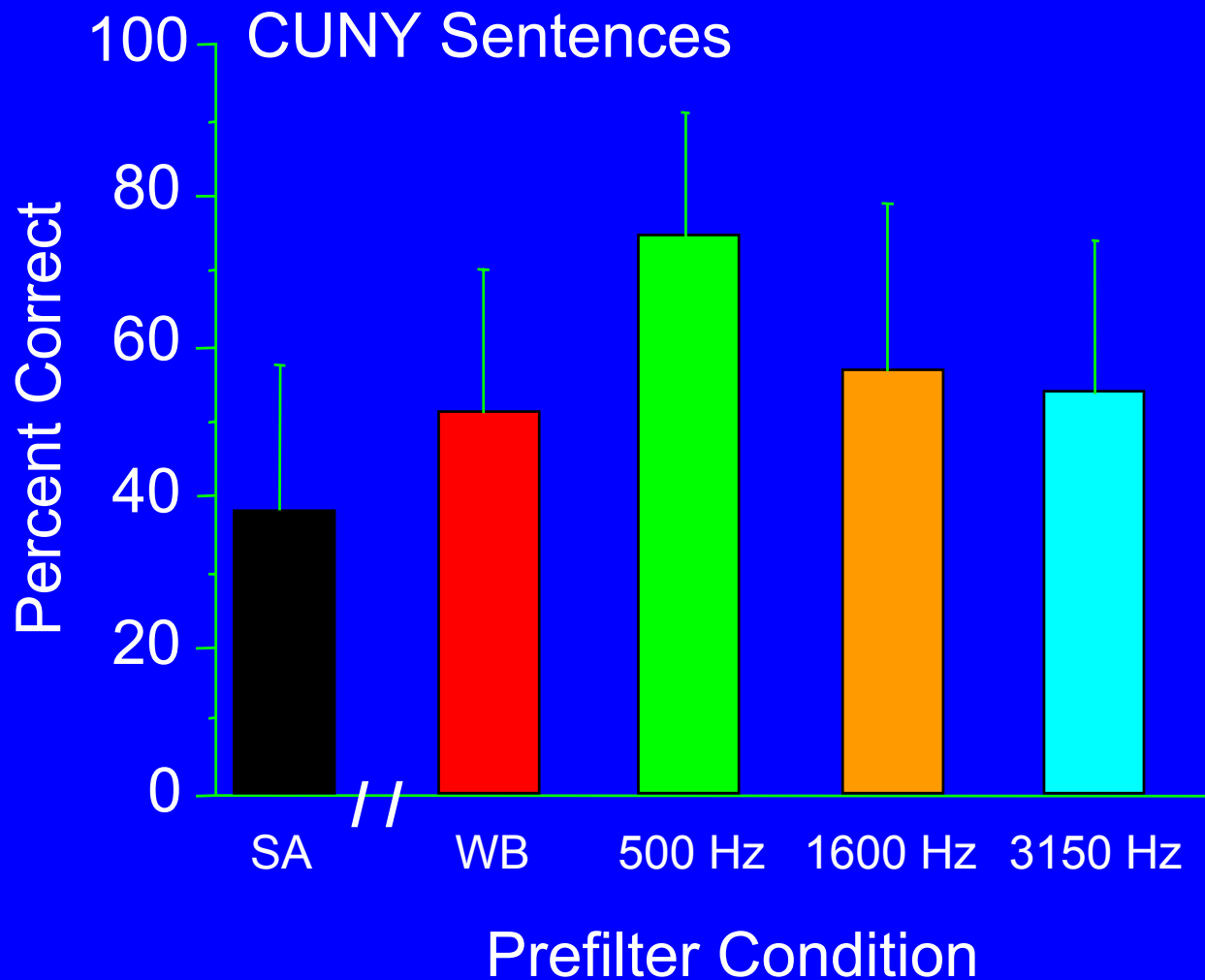
- AV Benefit is determined primarily by redundancy between acoustic and visual information
- Redundancy can be estimated by information transmission

Auditory-Visual Spectral Interactions: Consonants

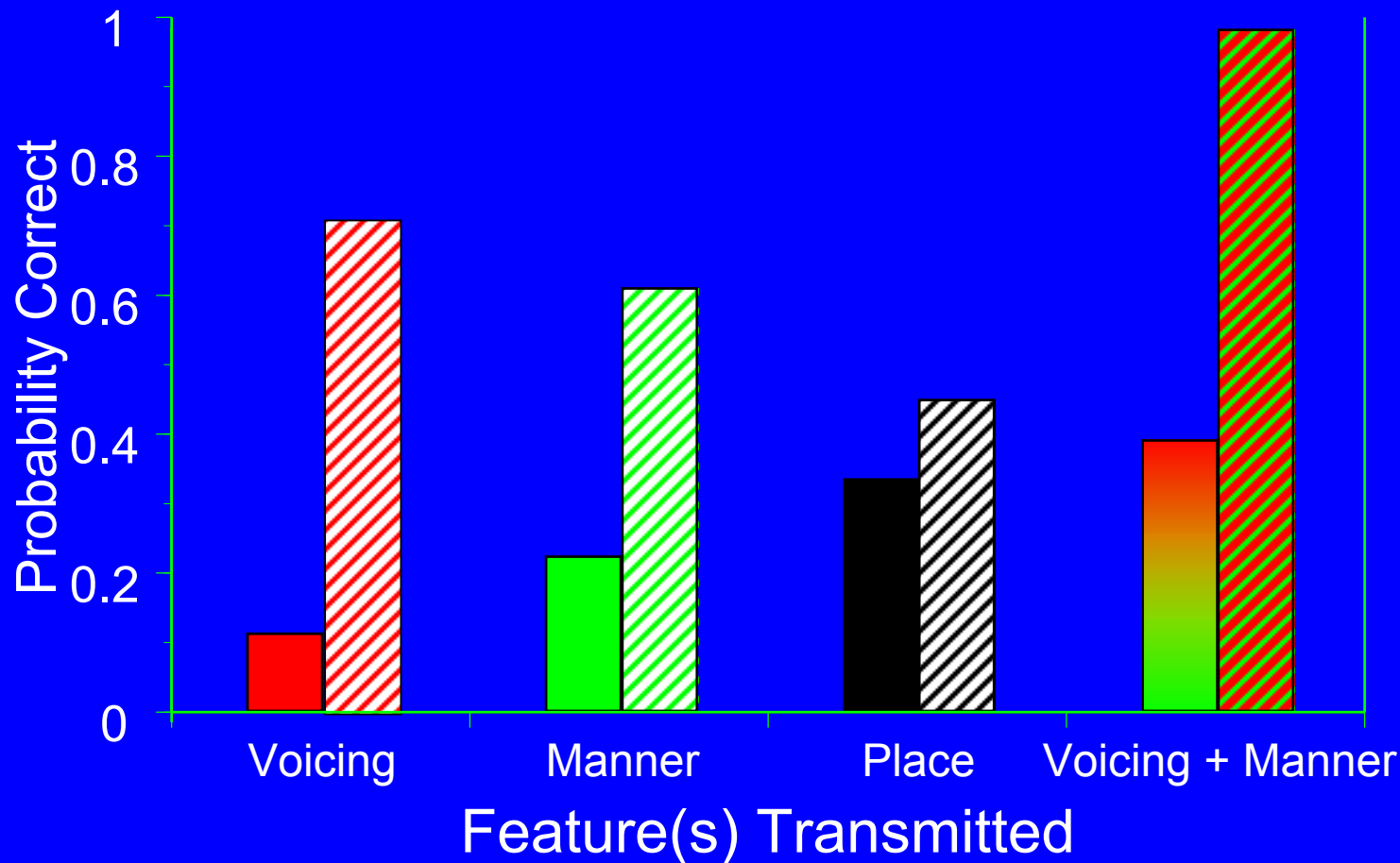





From Grant, K.W., and Walden, B.E. (1996). *J. Acoust. Soc. Am.* 100, 2415-2424.

Speechreading + Speech Envelope Bands



Redundancy Hypothesis – Modeling Results



   A

Auditory consonant recognition based on perfect transmission of indicated feature. Responses within each feature category were uniformly distributed.

   PRE

Predicted AV consonant recognition based on PRE model of integration (Braida, 1991).

Spectral Interactions - Summary

- Speechreading provides information mostly about place-of-articulation

Spectral Interactions - Summary

- Speechreading provides information mostly about place-of-articulation
- Auditory-visual speech recognition is determined primarily by complementary cues between visual and auditory modalities

Spectral Interactions - Summary

- Speechreading provides information mostly about place-of-articulation
- Auditory-visual speech recognition is determined primarily by complementary cues between visual and auditory modalities
- The most intelligible auditory speech signals do not necessarily result in the most intelligible auditory-visual speech signal

Spectral Interactions - Summary

- Speechreading provides information mostly about place-of-articulation
- Auditory-visual speech recognition is determined primarily by complementary cues between visual and auditory modalities
- The most intelligible auditory speech signals do not necessarily result in the most intelligible auditory-visual speech signal
- Acoustic cues for voicing and manner-or articulation are the best supplements to speechreading

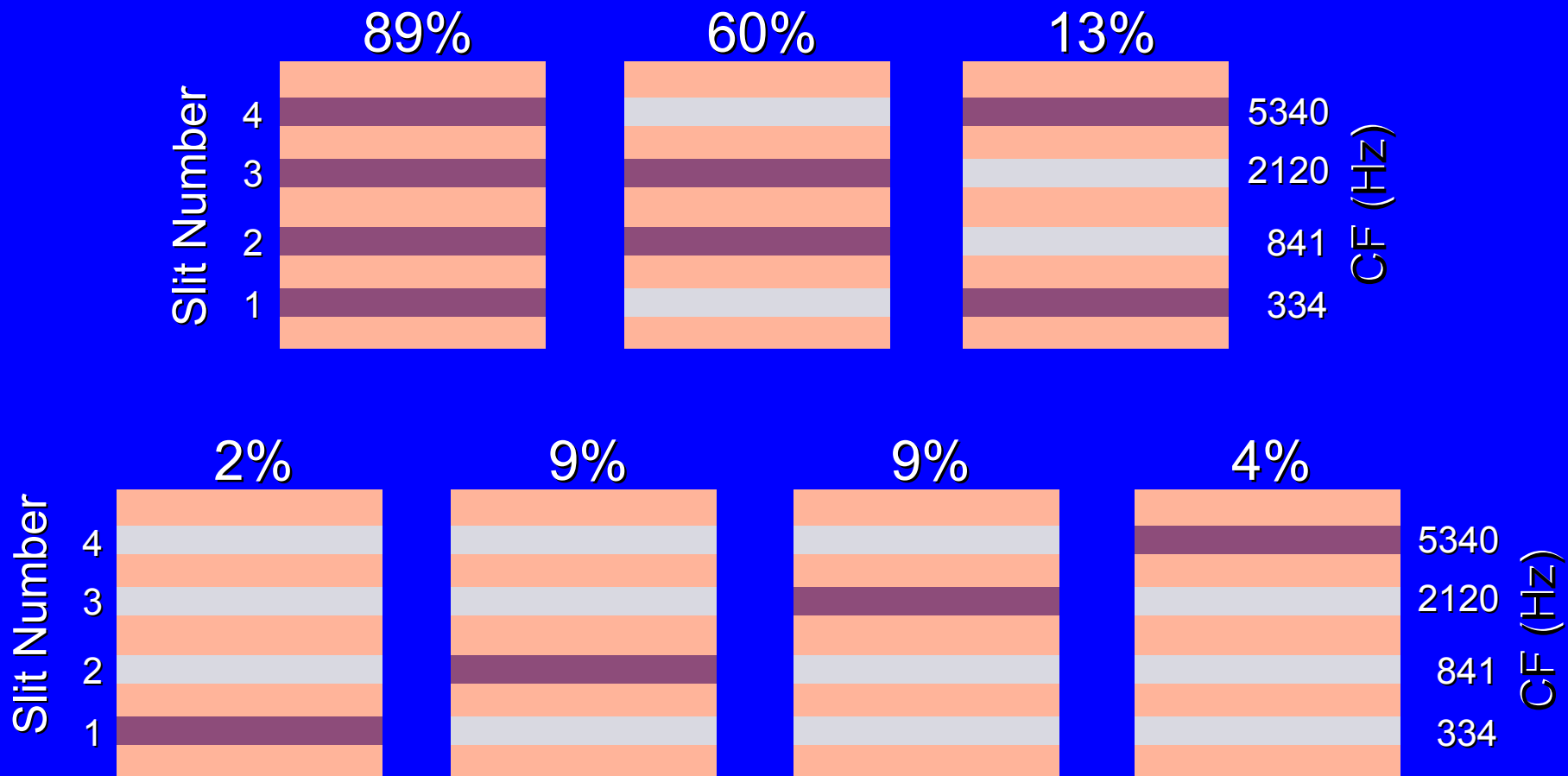
Spectral Interactions - Summary

- Speechreading provides information mostly about place-of-articulation
- Auditory-visual speech recognition is determined primarily by complementary cues between visual and auditory modalities
- The most intelligible auditory speech signals do not necessarily result in the most intelligible auditory-visual speech signal
- Acoustic cues for voicing and manner-or articulation are the best supplements to speechreading
- These cues tend to be low frequency

Temporal Window for A and AV Integration

*AUDIO-ALONE
EXPERIMENTS*

Word Intelligibility - Single and Multiple Slits

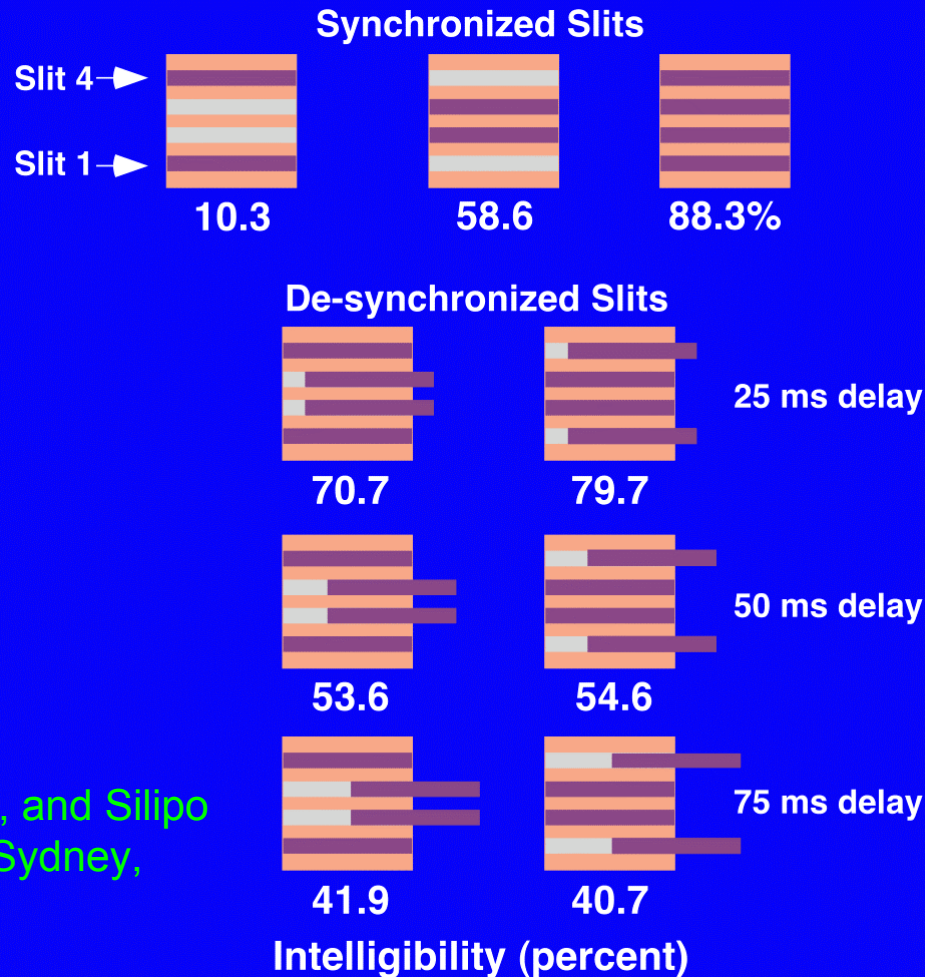


From Greenberg, Arai, and Silipo (1998). Proc. ICSLP, Sydney, Dec. 1-4.

Slit Asynchrony Affects Intelligibility

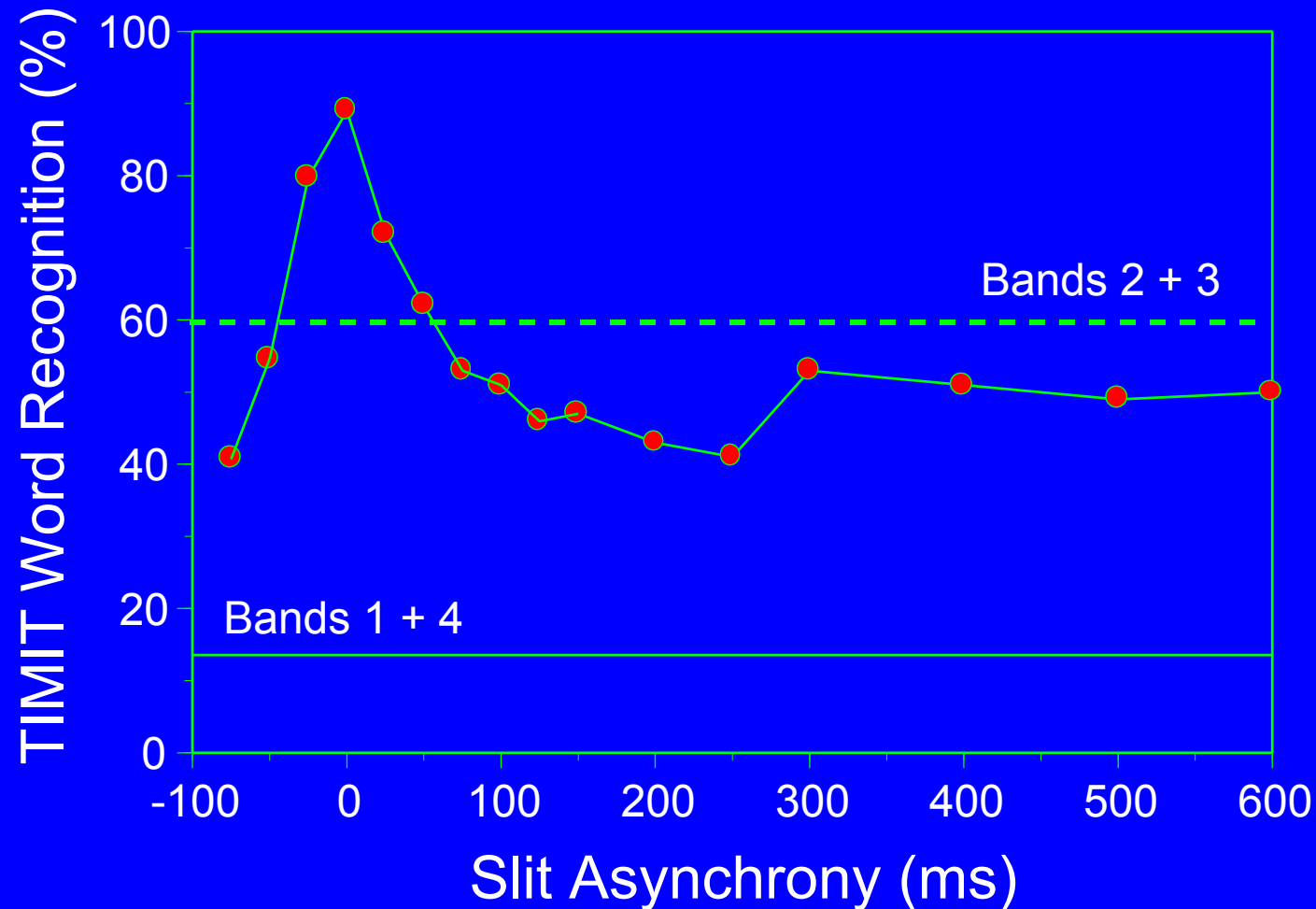
Desynchronizing the slits by more than 25 ms results in a significant decline in intelligibility

The effect of asynchrony on intelligibility is relatively symmetrical



From Greenberg, Arai, and Silipo (1998). Proc. ICSLP, Sydney, Dec. 1-4.

Cross-Spectral Temporal Asynchrony Effects



From Greenberg, Arai, and Silipo (1998). Proc. ICSLP, Sydney, Dec. 1-4.

*AUDITORY-VISUAL
EXPERIMENTS*

Auditory-Visual Tasks

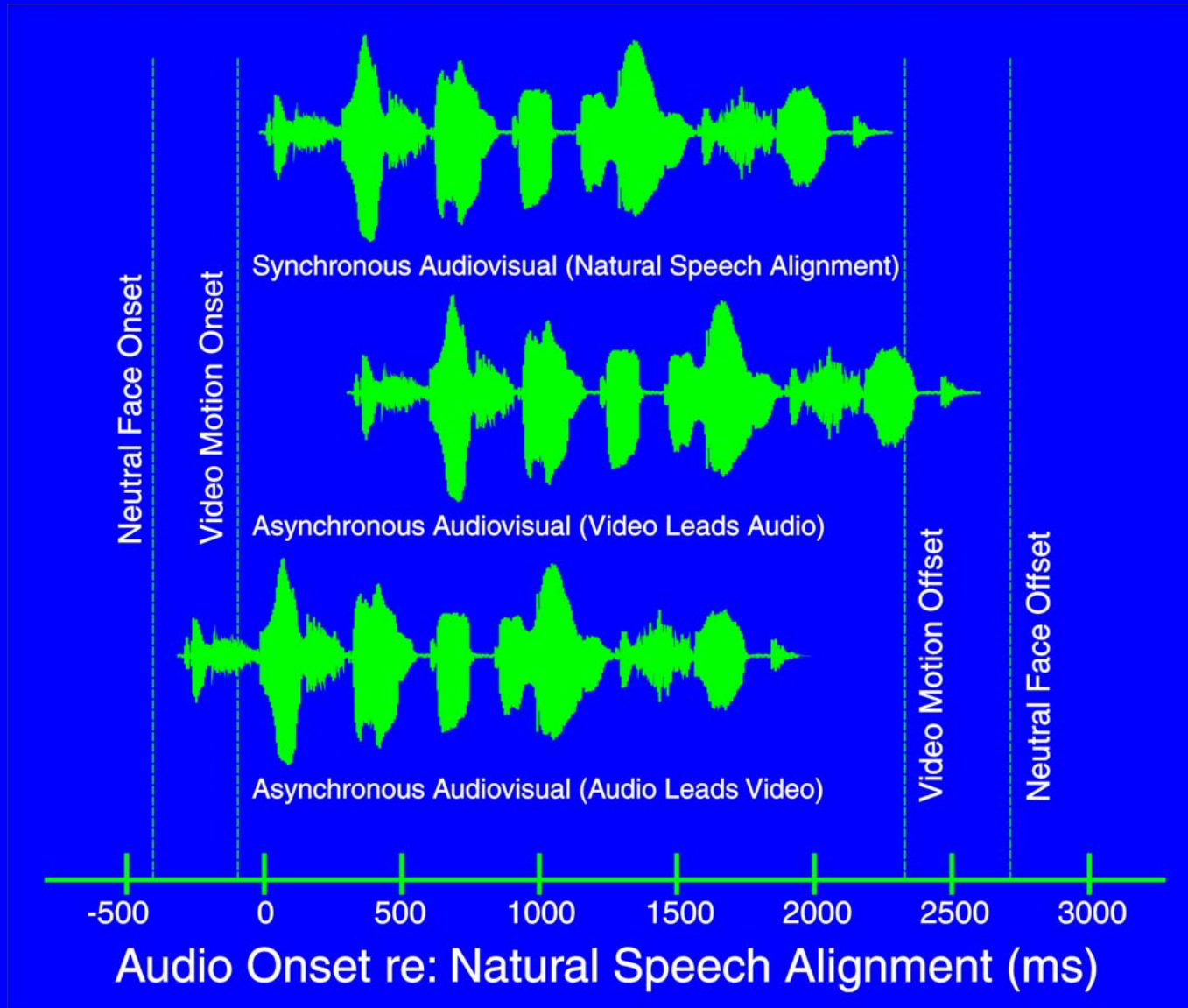
IEEE Sentences

- Recognition of key words
 - Audio slits 1 + 4
 - Video presented at various temporal asynchronies

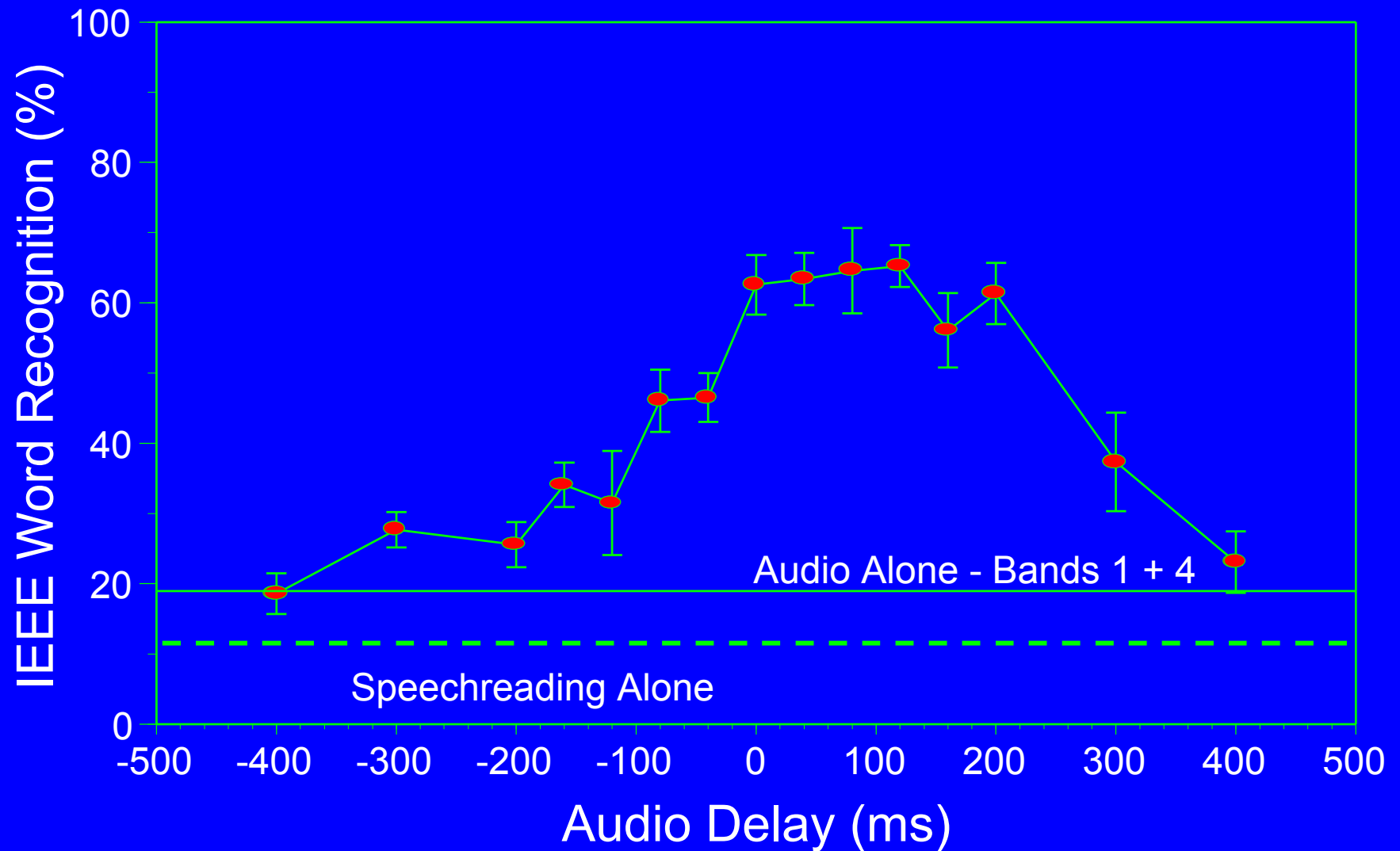
CV Syllables

- Recognition of McGurk pairs
 - Audio /pa/, /ba/, /ta/, /da/
 - Video /ka/, /ga/, /ta/, /da/
- Synchrony identification and discrimination
 - Yes/No single interval simultaneity judgments
 - congruent versus incongruent tokens

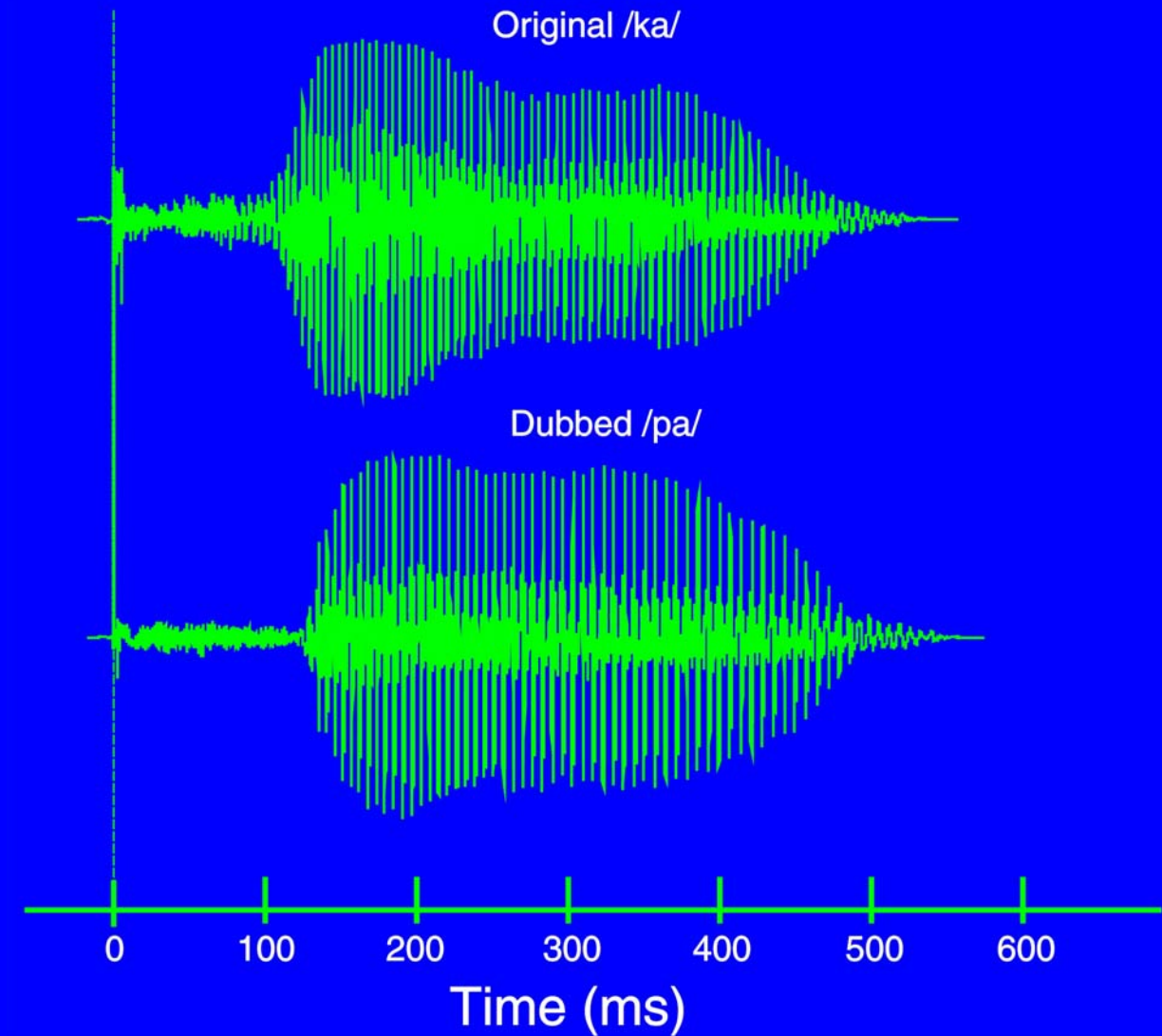
Auditory-Visual Asynchrony - Paradigm



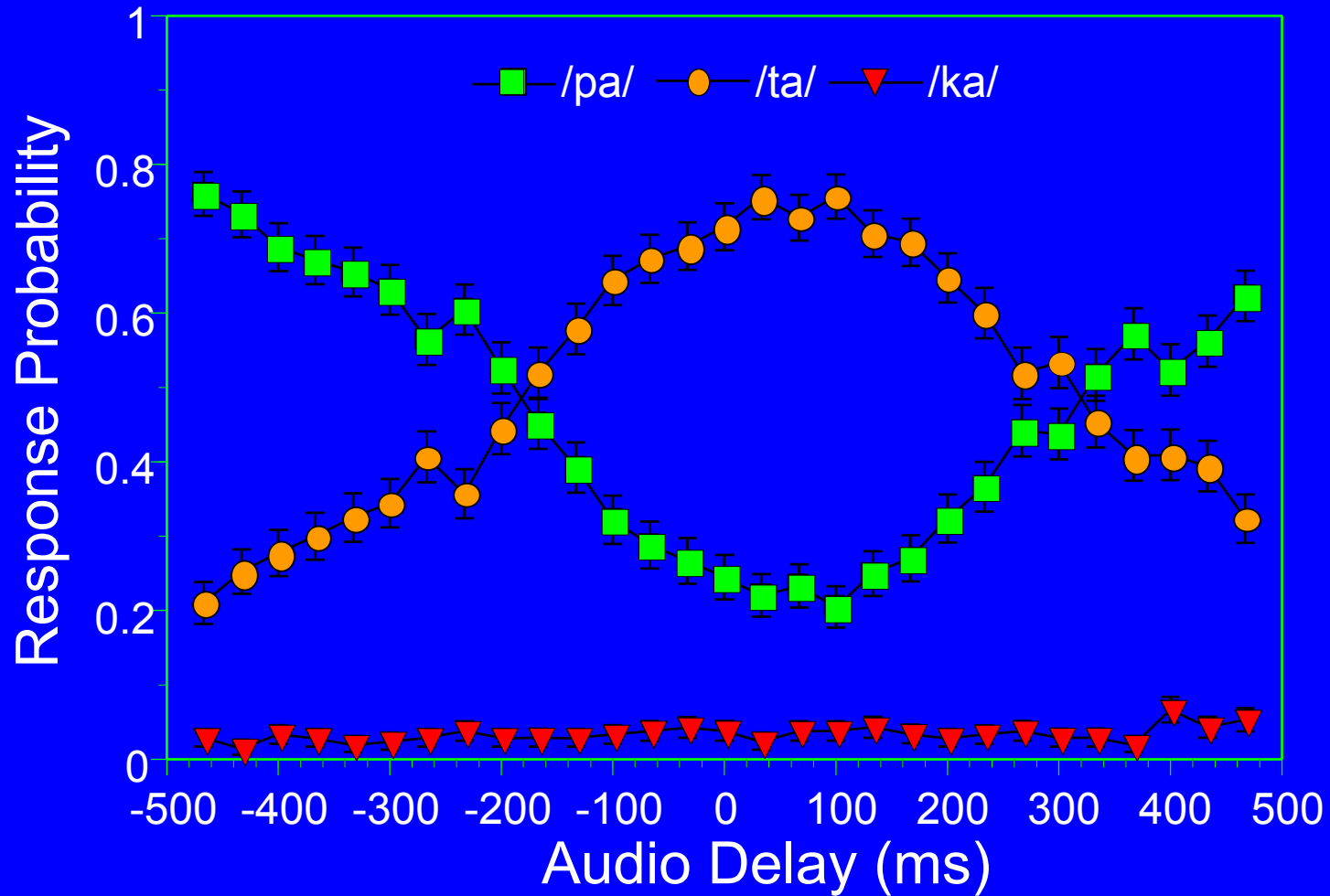
Cross-Modality Temporal Asynchrony Effects: Sentences



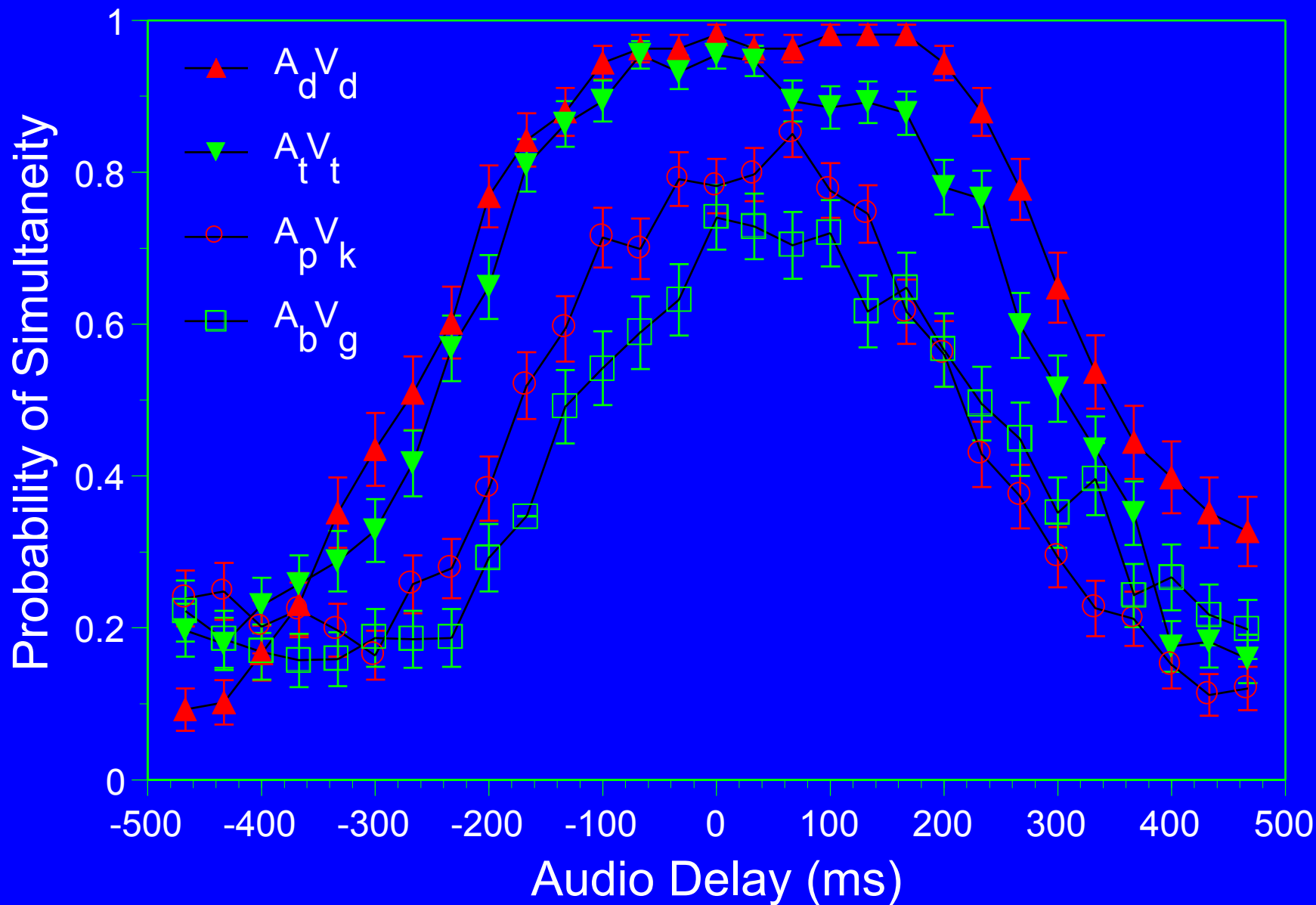
McGurk Synchrony Paradigm



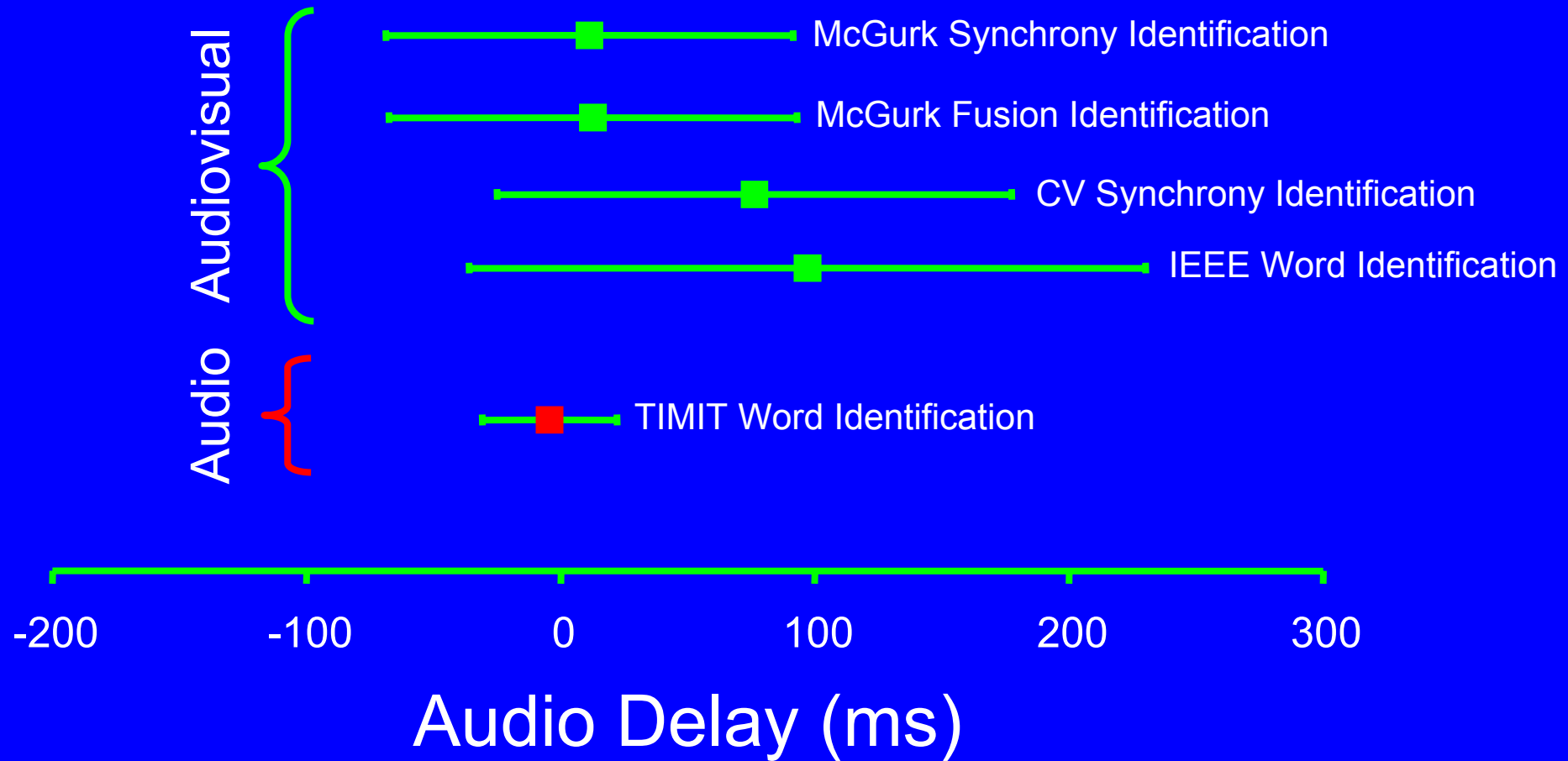
Temporal Integration in the McGurk Effect



Synchrony Identification - Natural vs. McGurk AV Tokens



Temporal Window of Integration



Spectro-Temporal Integration: Summary

Within Modality (Cross- Spectral Auditory Integration)

- TWI is symmetrical
- TWI roughly 20-40 ms (phoneme?)

Across Modality (Cross-Modal AV Integration)

- TWI is highly asymmetrical favoring visual leads
- TWI is roughly 160-250 ms (syllable?)
- TWI for Incongruent CV's (McGurk Stimuli) is not as wide as TWI for natural congruent CV's

Auditory-Visual Speech Perception Laboratory



Walter Reed Army Medical Center
Army Audiology and Speech Center
Washington, DC USA

<http://www.wramc.amedd.army.mil/departments/aasc/avlab>
grant@tidalwave.net