Discrimination of Auditory-Visual Synchrony

Ken W. Grant

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

Virginie van Wassenhove

David Poeppel

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

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

grant@tidalwave.net

Acknowledgments

Collaborations:

Steven Greenberg - The Speech Institute, Oakland, CA

Funding:

- NIH Grant: DC 00792-01AI – Auditory Supplements to Speechreading
- NSF Grant (subcontract): SBR 9720398 – Learning and Intelligent Systems Initiative
- DARPA Grant (subcontract): ONR Award N000140210571 – Augmented Cognition

GOALS

Determine thresholds for temporal synchrony in auditory visual (AV) speech processing for different spectral channels (spectro-temporal, cross-modality integration)

Determine thresholds for temporal synchrony for different spectral channels in auditory speech processing (spectro-temporal, within-modality integration)

Motivation

Auditory-visual integration:

1. does not require precise temporal alignment – temporal disparities up to 200-350 ms appear to be well tolerated (Dixon and Spitz, 1980; McGrath and Summerfield, 1985; Massaro et al., 1996)
2. Visual leads tolerated much more than visual lags

Motivation (continued)

Auditory integration (across spectral channels):

1. highly tolerant to cross-spectral asynchrony *if* there are a large number of redundant channels (Arai and Greenberg, 1998)
2. with few spectral channels, sensitivity to cross-spectral asynchrony is high (Silipo et al., 1999)

Motivation (continued)

Determining the temporal window of integration (TWI) for both AV (cross modality) and A (cross spectral) speech processing has implications for:

Motivation (continued)

Determining the temporal window of integration (TWI) for both AV (cross modality) and A (cross spectral) speech processing has implications for:

1. models of integration and speech perception

Motivation (continued)

Determining the temporal window of integration (TWI) for both AV (cross modality) and A (cross spectral) speech processing has implications for:

1. models of integration and speech perception
2. constrain neurophysiological mechanisms that underlie unisensory and multisensory integrative processes

Motivation (continued)

Determining the temporal window of integration (TWI) for both AV (cross modality) and A (cross spectral) speech processing has implications for:

1. models of integration and speech perception
2. constrain neurophysiological mechanisms that underlie unisensory and multisensory integrative processes
3. defining temporal limits for signal processing operations

Background

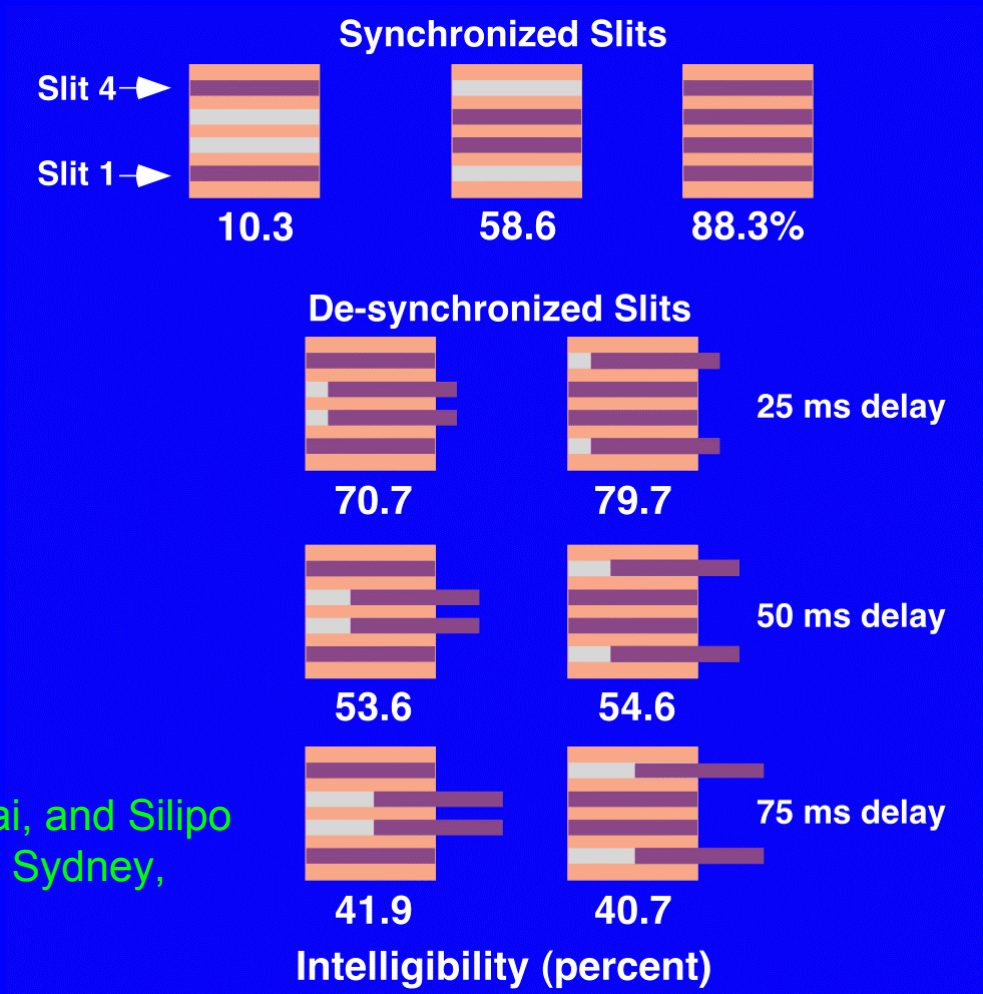
Brief review of auditory and auditory-visual temporal effects on speech recognition

AUDIO-ALONE EXPERIMENTS

Slit Asynchrony Affects Intelligibility

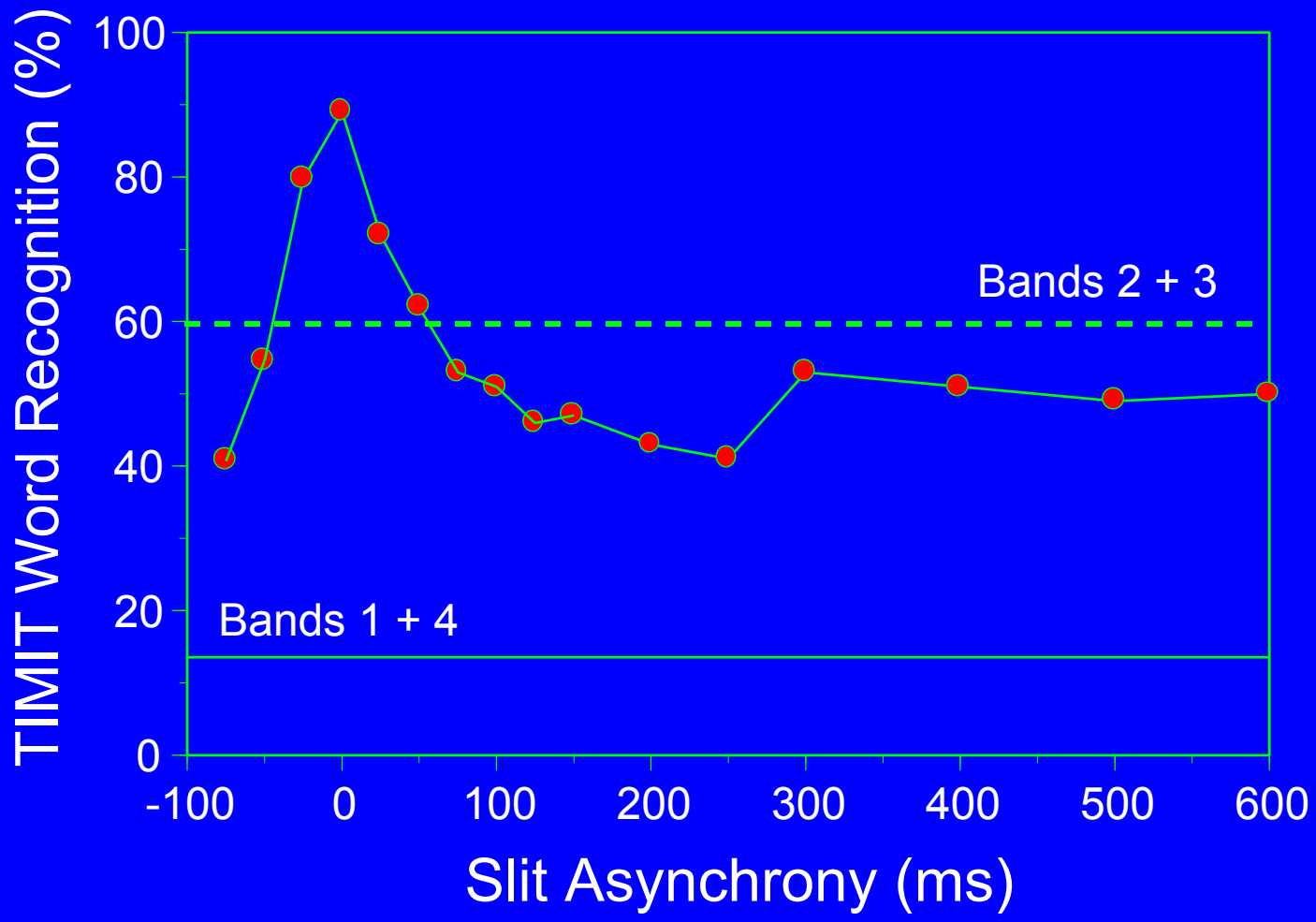
Desynchronizing the slits by more than 25 ms results in a significant decline in intelligibility

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

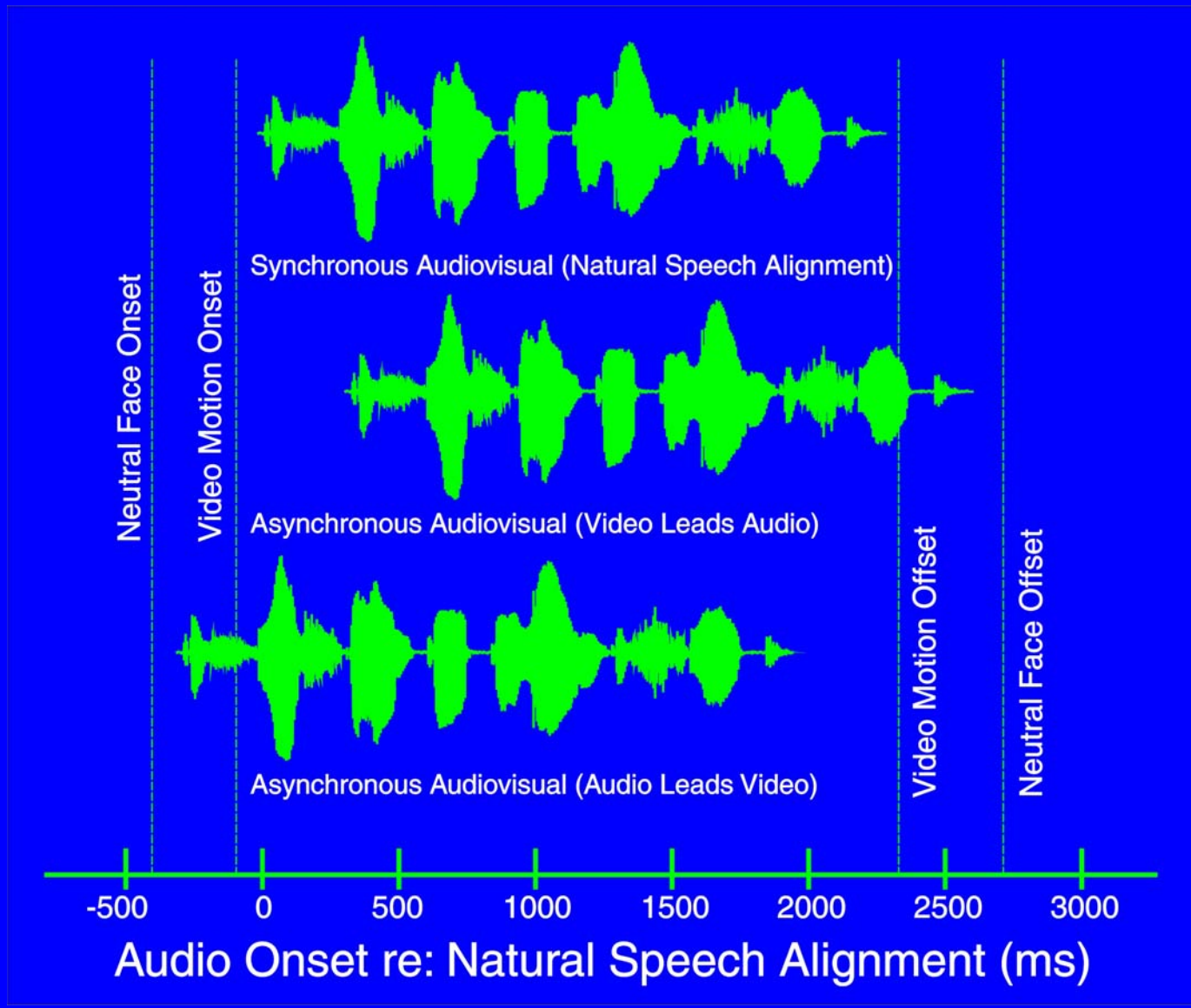
IEEE Sentences

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

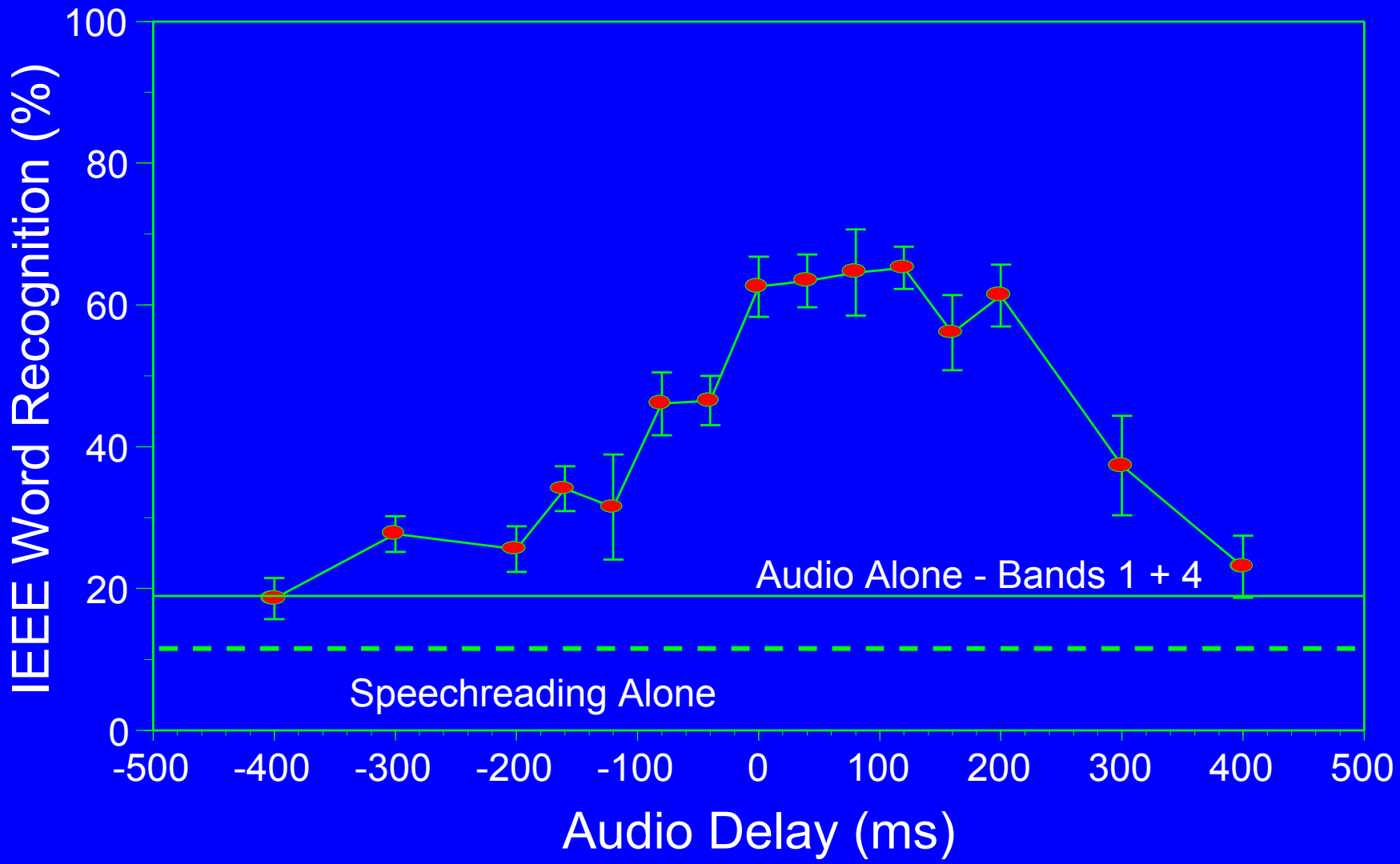
CV Syllables

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

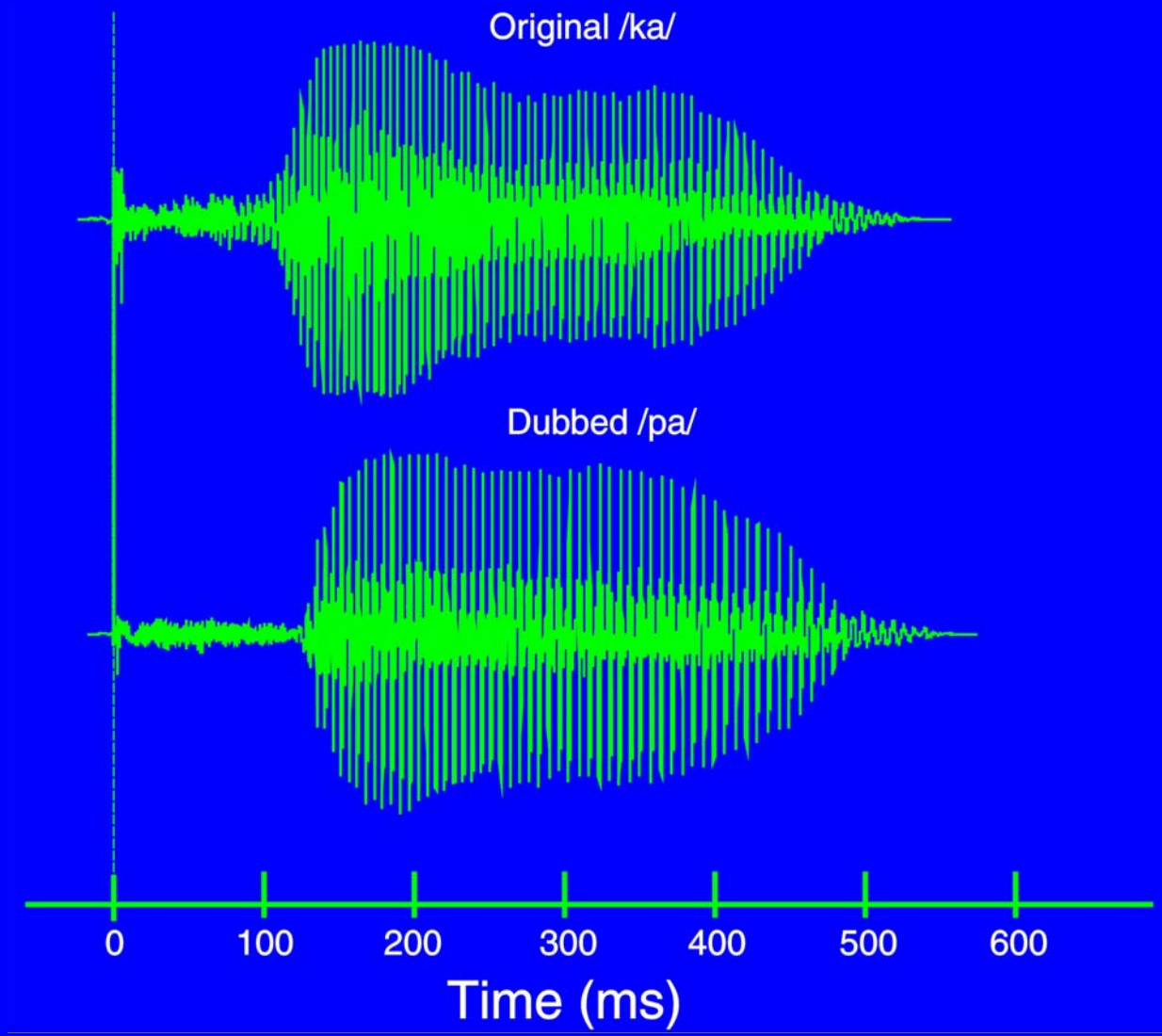
Auditory-Visual Asynchrony - Paradigm



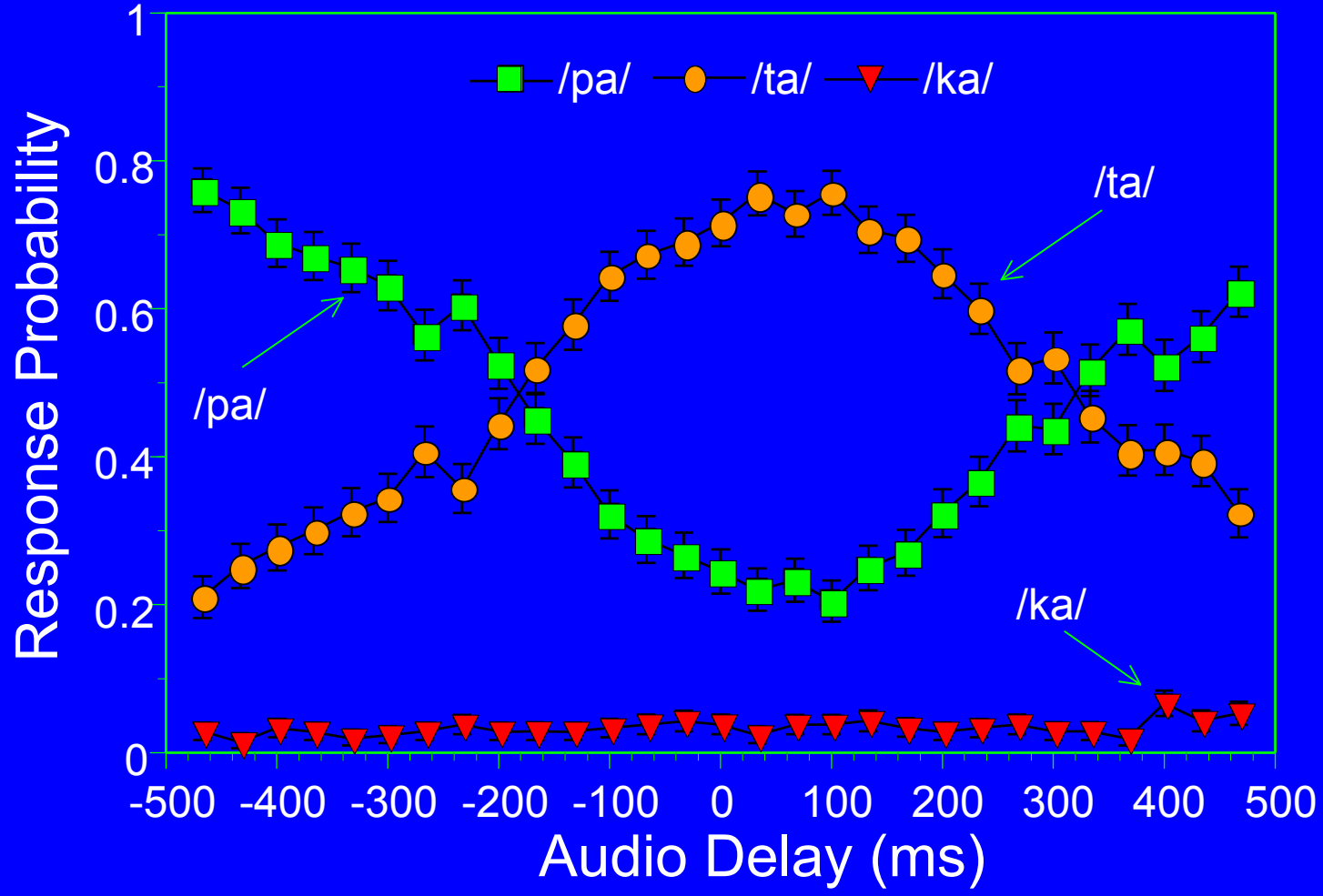
Cross-Modality Temporal Asynchrony Effects: Sentences



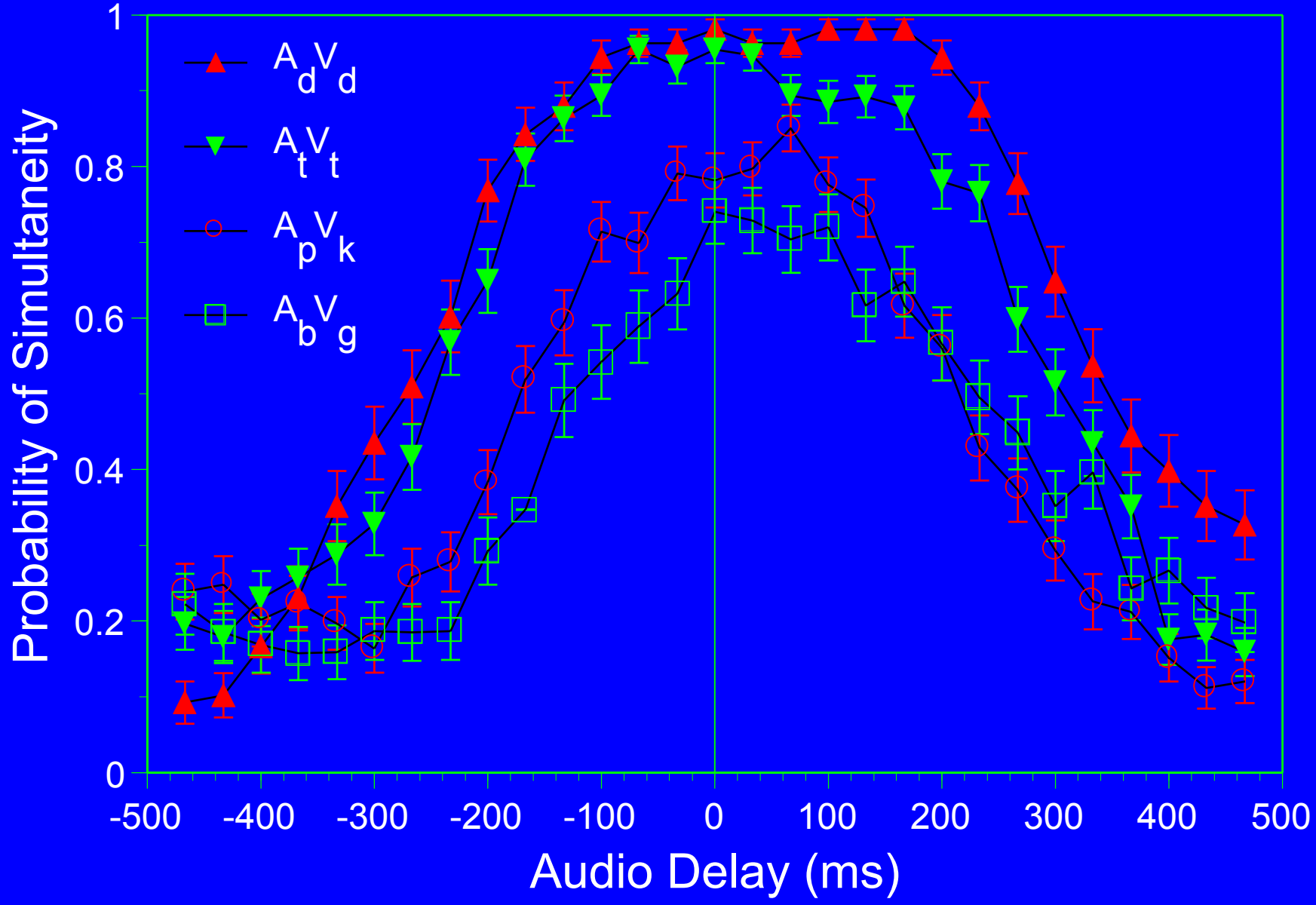
McGurk Synchrony Paradigm



Temporal Integration in the McGurk Effect



Synchrony Identification - Natural vs. McGurk AV Tokens

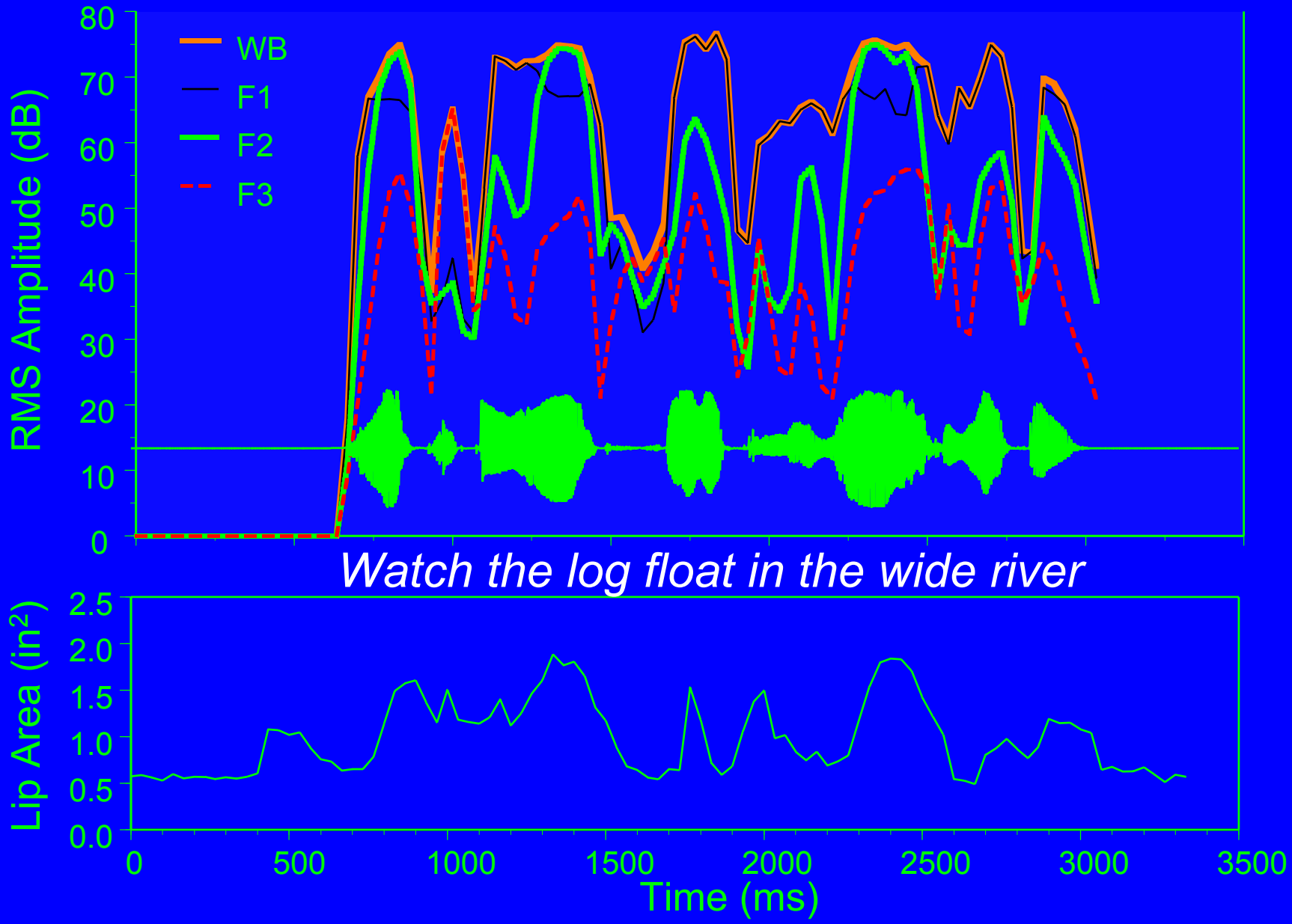


AV Spectro-Temporal Coherence

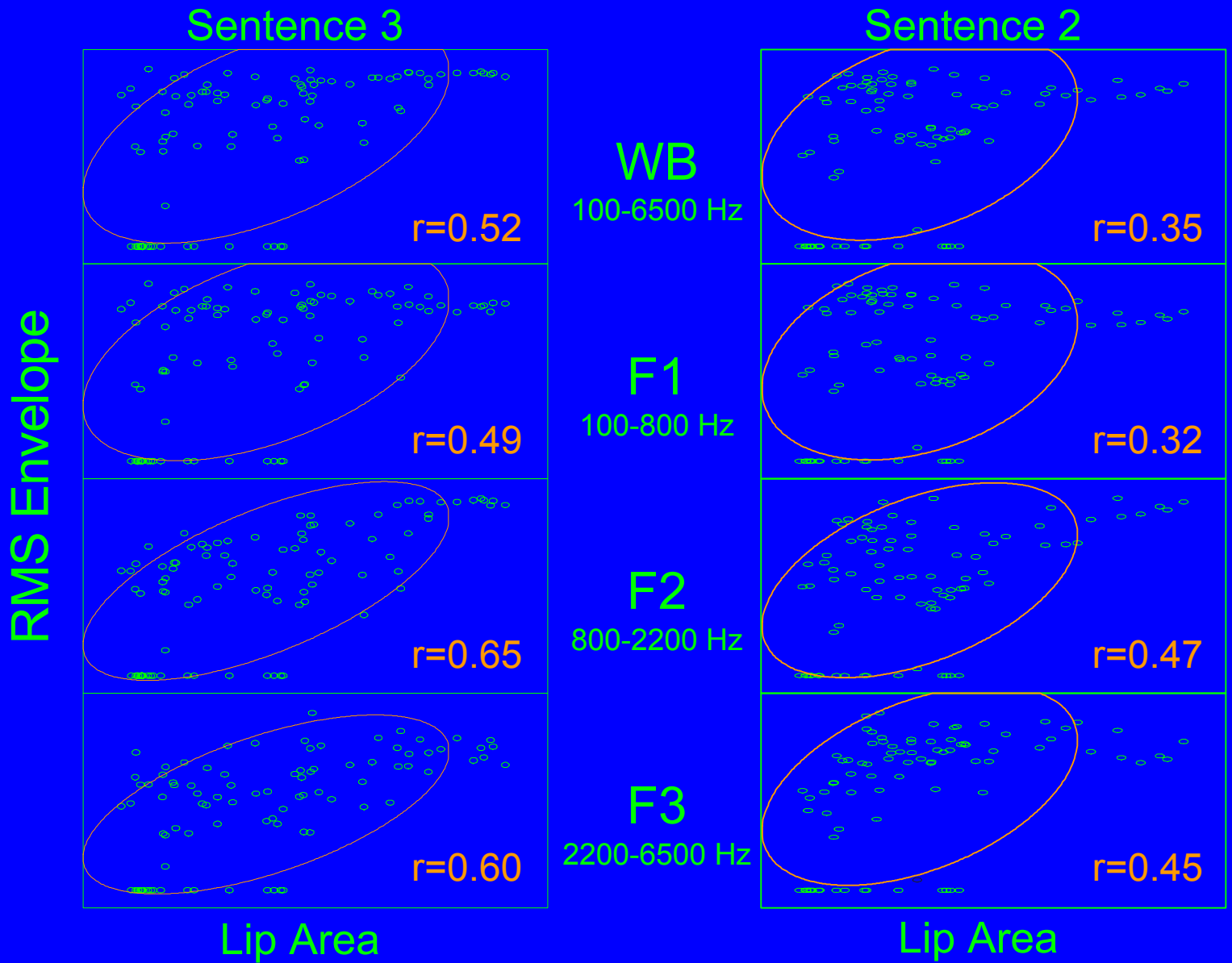
Visible articulatory kinematics are correlated with acoustic envelope (Grant and Seitz, 2000)

Degree of correlation depends on the spectral band (highest correlation found for mid-frequency bands in the F2-F3 region

Acoustic Envelope and Lip Area Functions



Cross Modality Correlation - Lip Area versus Amplitude Envelope



Current AV Experiment

Sentence materials

2IFC adaptive tracking procedure with feedback

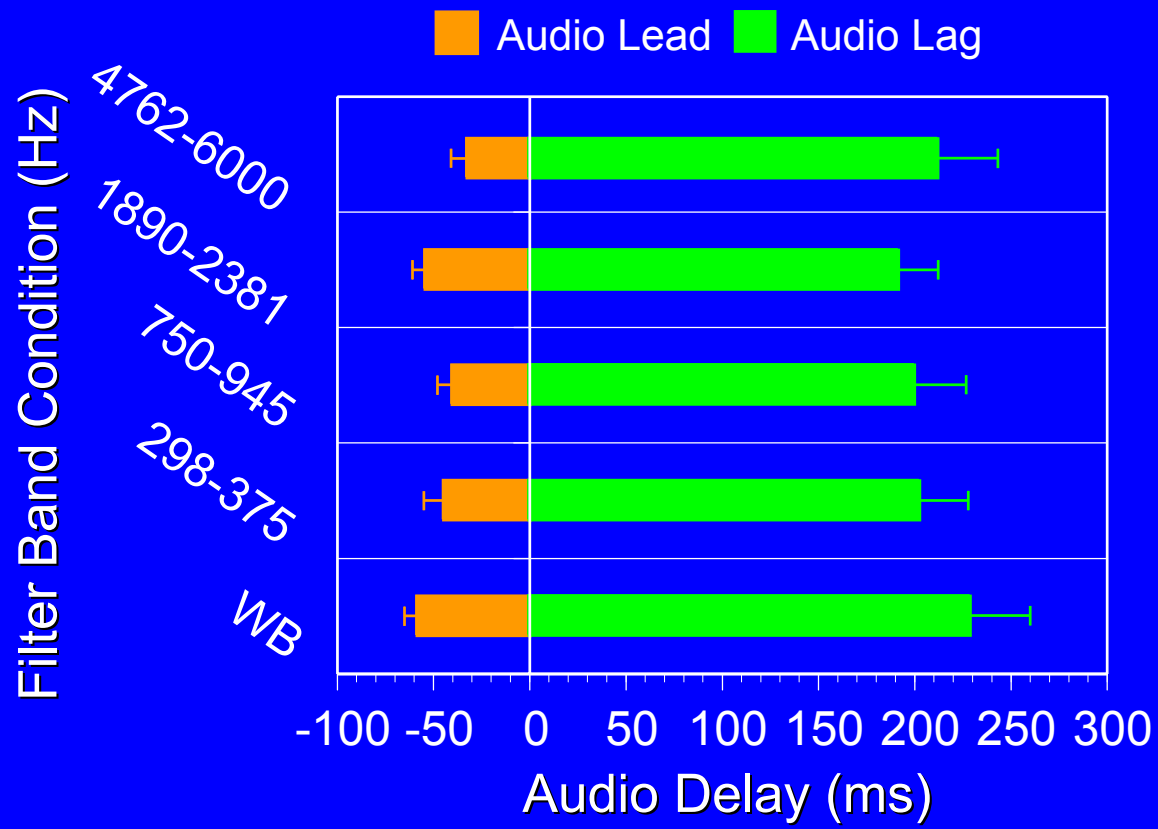
Thresholds for AV asynchrony (leading and lagging)

5 different filter band conditions

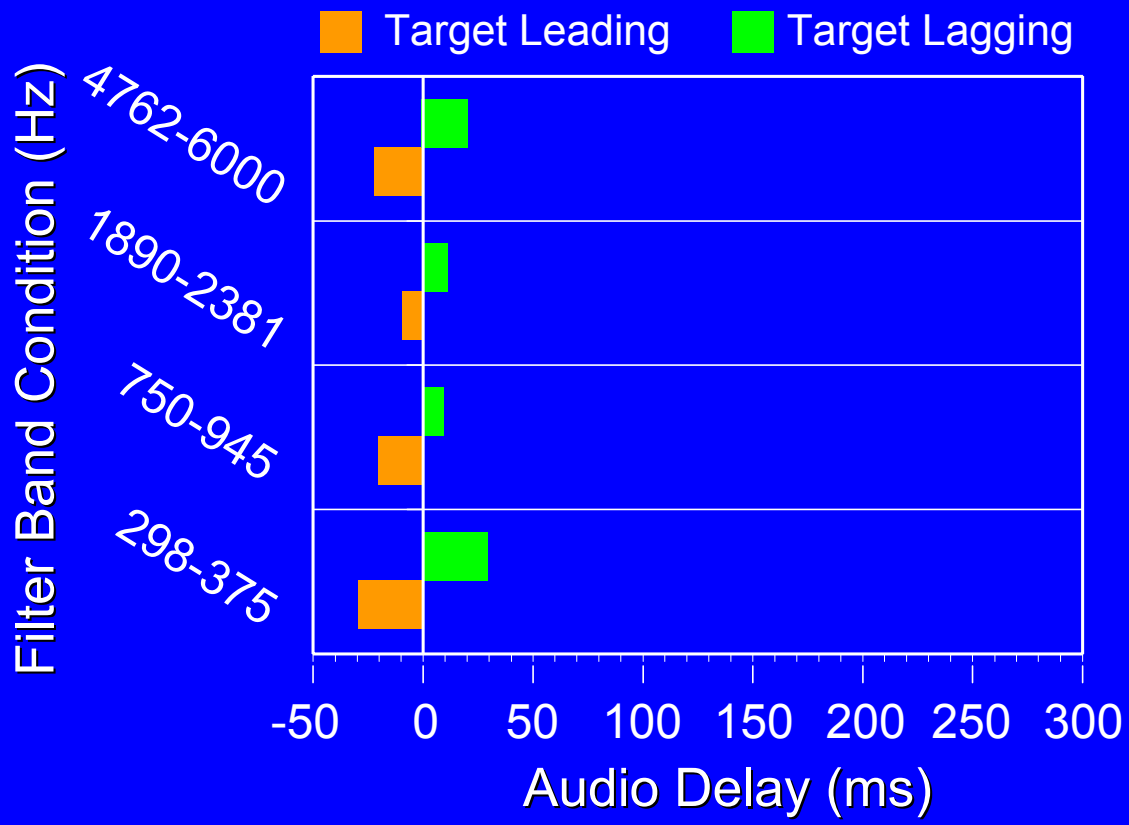
- Wideband
- 298-375 Hz
- 750-945 Hz
- 1890-2381 Hz (highest correlation)
- 4762-6000 Hz

4 adult normal-hearing subjects (35-49 years)

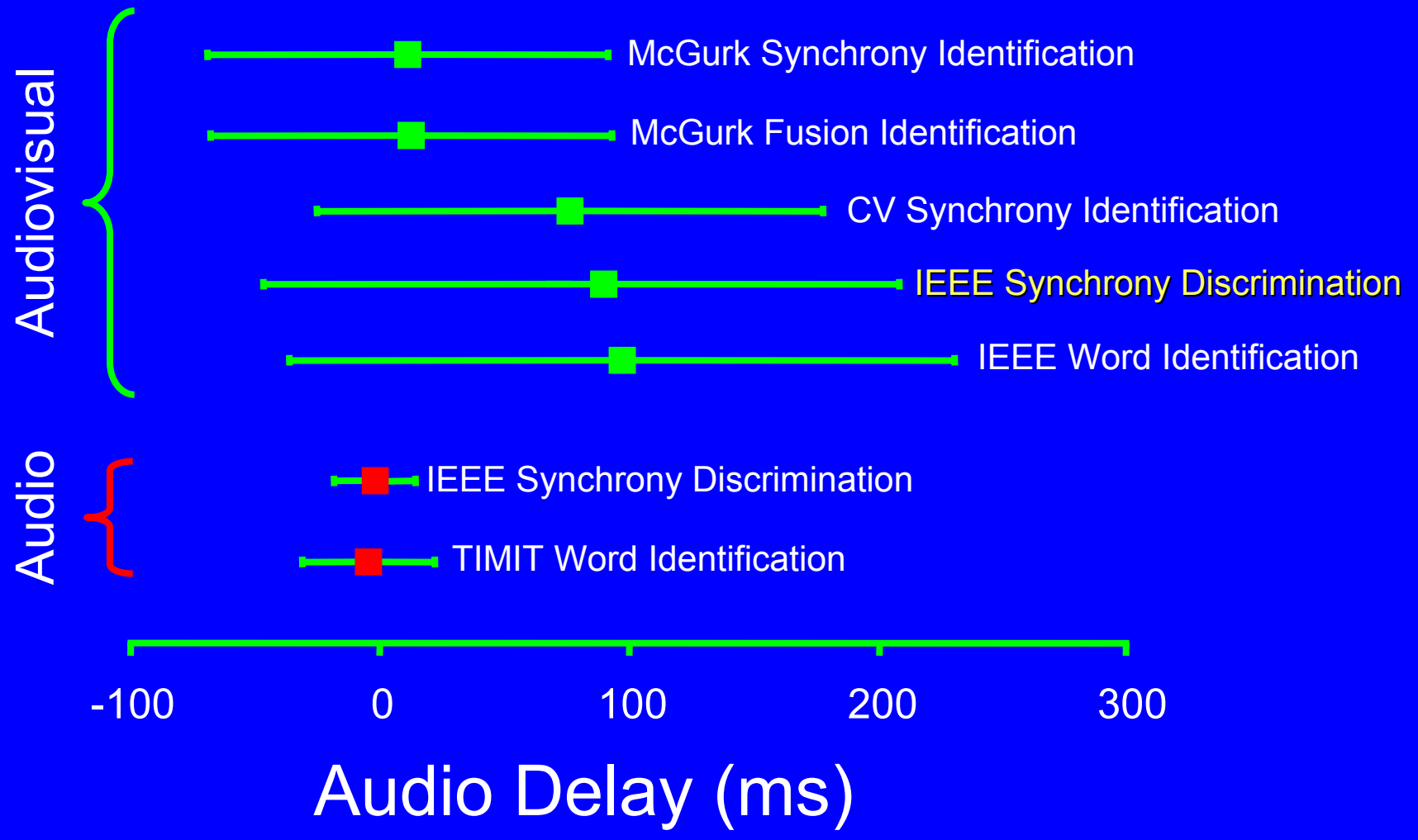
AV Spectro-Temporal Synchrony Discrimination



Auditory Spectro-Temporal Synchrony Discrimination



Temporal Window of Integration



Spectro-Temporal Integration: Summary

Threshold for Synchrony Discrimination Determines Cross-Modal and Cross-Spectral TWI

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

Within Modality (Cross- Spectral Auditory Integration)

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

Auditory-Visual Speech Perception Laboratory

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

grant@tidalwave.net