Spectro-Temporal Interactions in Auditory-Visual Perception:

How the Eyes Modulate What the Ears Hear

Ken W. Grant

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

Virginie van Wassenhove

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

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

grant@tidalwave.net

Primary Goal

- In what ways can visual cues influence auditory processing of acoustic events?
  - Focus on speech perception

    BUT

  - Visual cues also can influence non-speech event perception
    - Caveat: AV interactions for non-speech events may be smaller than for speech events

Organizational Framework

Visual cues alter the perception of acoustic events at all levels:

- Event Detection
- Event Localization (ventriloquism)
- Event Identification
- Event Quality

Pragmatic Concerns

Given that visual cues impact on many dimensions of sound perception:

- **How might this information change the way acoustic engineers design concert halls?**
  - less emphasis on sound simulations which omit visual cues
  - greater understanding of the perceptual consequences of multimodal input (e.g., loudness constancy)
  - may lead to new designs which explicitly exploit these effects to enrich the audience experience.

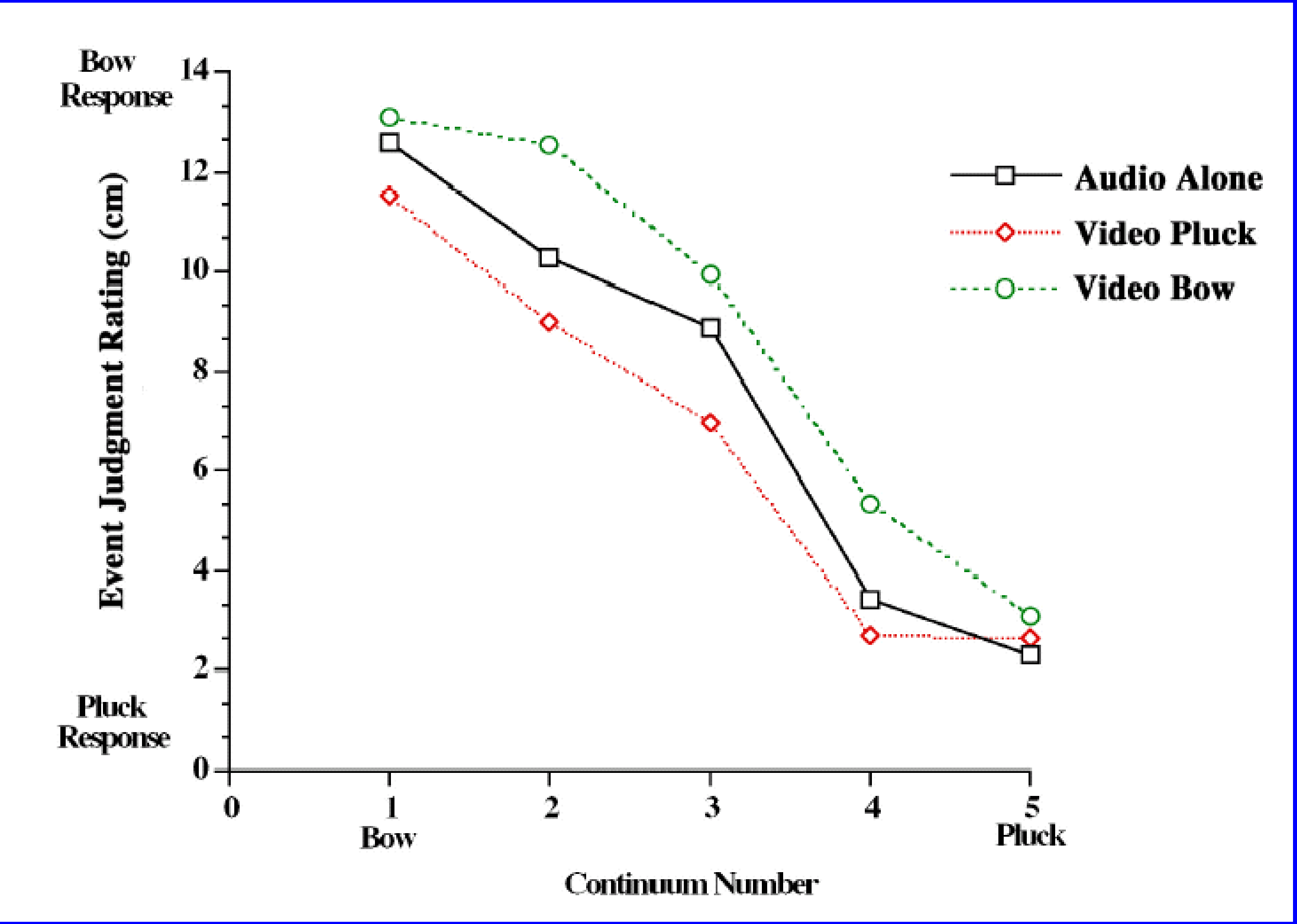
AV Interactions for Non-Speech Events

Perception of Musical Identity

H.M. Saldaña, and L.D. Rosenblum (1993). Visual influences on auditory pluck and bow judgments. Perception and Psychophysics, 54, 406-416

- Synthesize a continuum from bowed sounds to plucked sounds
- Subjects rate sounds (bow or pluck) under A and AV conditions
- Visual cues consist of a movie of a hand either plucking or bowing a string

Pluck versus Bow



from Saldaña and Rosenblum (1993)

Demonstration

Special Thanks to Marcelo Wanderly and Bradley Vines

Perception of Musical Tension and Phrasing

B. Vines, M. M. Wanderley, C. Krumhansl, R. Nuzzo, and D. Levitin. Performance Gestures of Musicians: What Structural and Emotional Information do they Convey? In A. Camurri and G. Volpe (eds.) Gesture-Based Communication in Human-Computer Interaction - 5th International Gesture Workshop, GW 2003, Genova, Italy. Springer-Verlag, 2004, pp. 468 - 478.

Procedures

- Continuous judgments sampled every 100 ms
- Tension

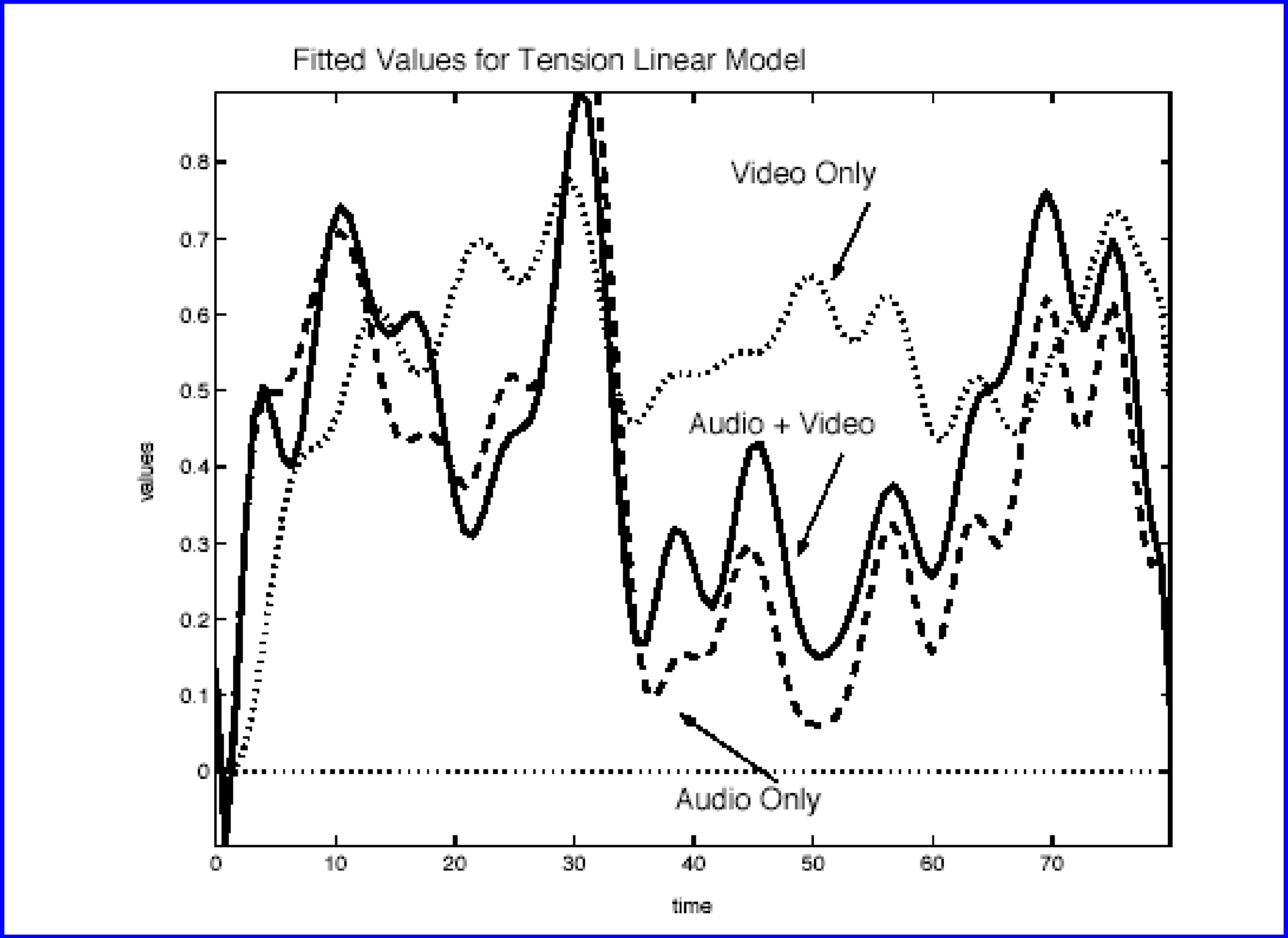
  Use the full range of the slider to express the TENSION you experience in the performance. Move the slider upward as the tension increases and downward as the tension decreases.
- Phrasing

  Use the full range of the slider to express the PHRASING you experience in the performance. Move the slider upward as a phrase is entered and downward as a phrase is exited. The slider should be near the top in the middle of a phrase and near the bottom between phrases.

Musical Demo

Solo clarinet performance. End of one musical phrase and beginning of next. Notice the contuation of movement at the end of the phrase, and more importantly, the early onset of movement signifying the beginning of the next phrase. This clear visual gesture prior to the onset of sound allows the audience to anticipate the onset of the phrase.

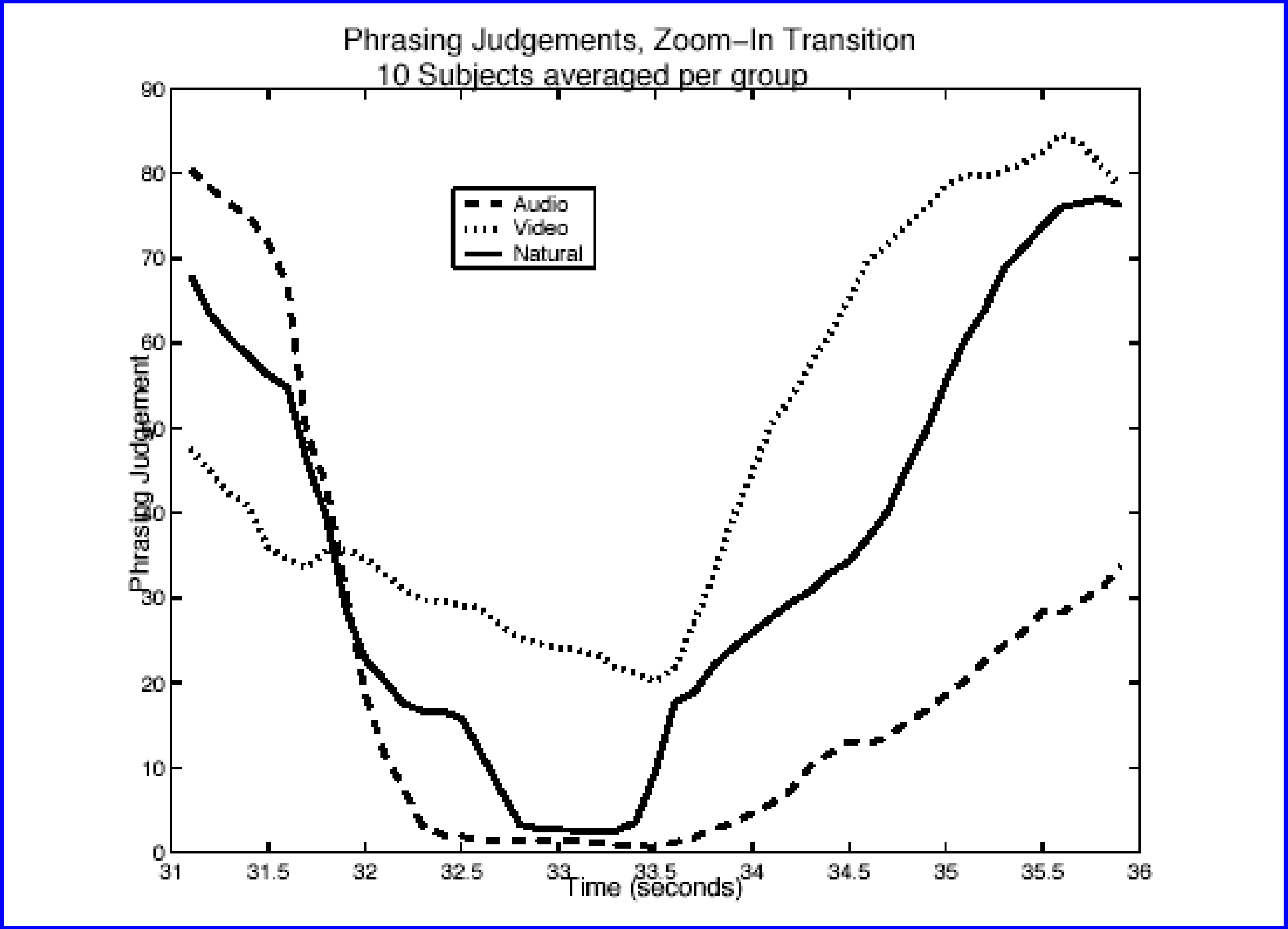
Perception of Musical Tension



from Vines et al. (2003)

Auditory only marking of tension is typically less than in the natural AV case

Perception of Musical Phrasing



from Vines et al. (2003)

Offset of initial phrase is clearly marked by either A or AV modes. Perception of the onset of following phrase is sluggish for A and slightly early in V.

Music Perception and Cognition

8th International Conference on Music Perception & Cognition
Northwestern University
August 3-7, 2004

B. Vines, R. Nuzzo, C. Krumhansl, & D. Levitin: Visual Music: The Perceptual Impact Of Seeing A Clarinetist (McGill University, Canada)

W.F. Thompson & F.A. Russo: Visual Influence on the Perceived Size of Sung Intervals (University of Toronto, Canada)

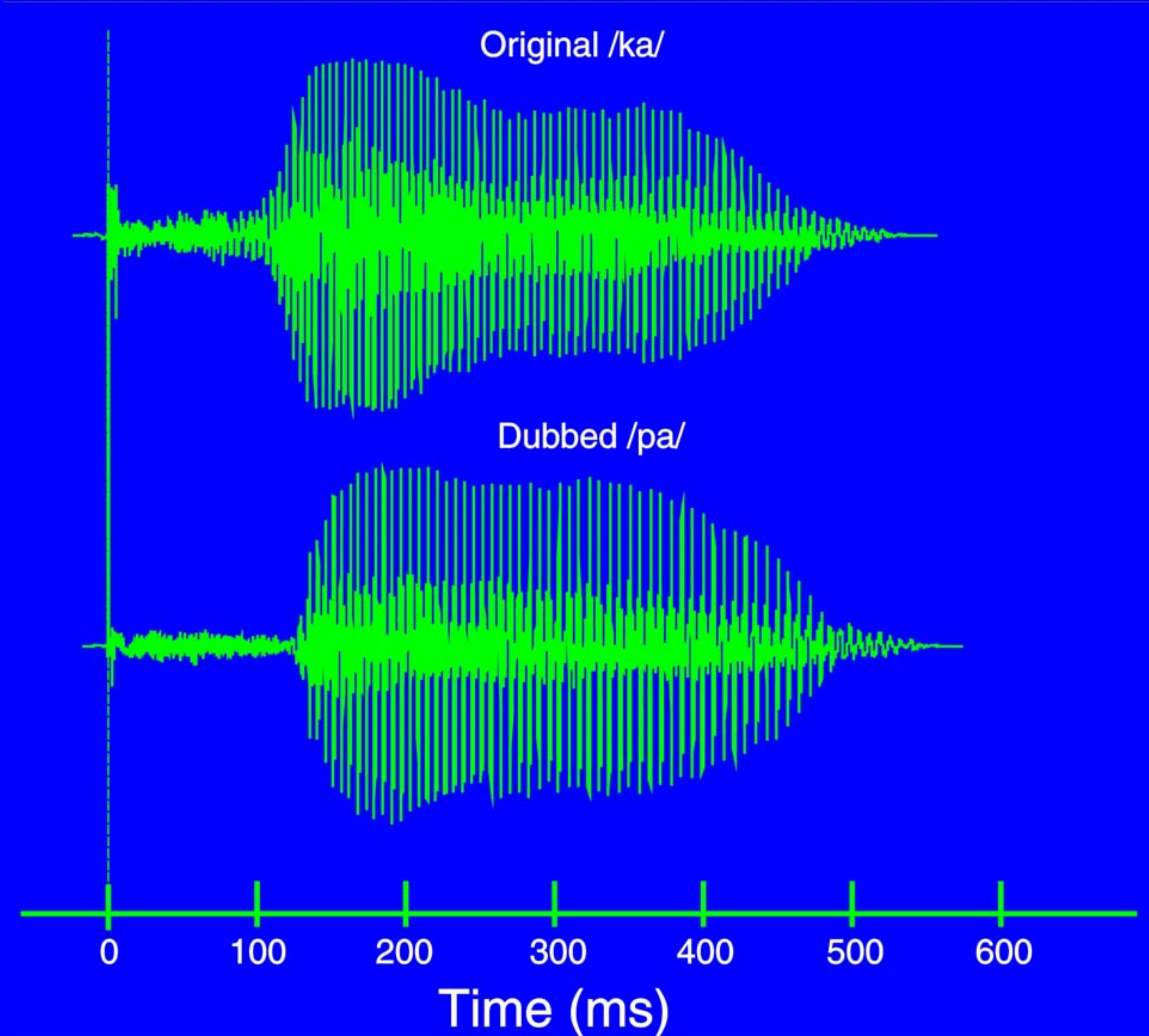
W.F. Thompson & F.A. Russo: Visual Influences on the Perception of Emotion in Music (University of Toronto, Canada)

K. Kallinen: Emotion Related Psychophysiological Responses to Listening to Music with Eyes-Open Versus Eyes-Closed: Electrodermal (EDA), Electrocardiac (ECG), and Electromyographic (EMG) Measures (Knowledge Media Laboratory, Helsinki School of Economics)

AV Interactions for Speech Events

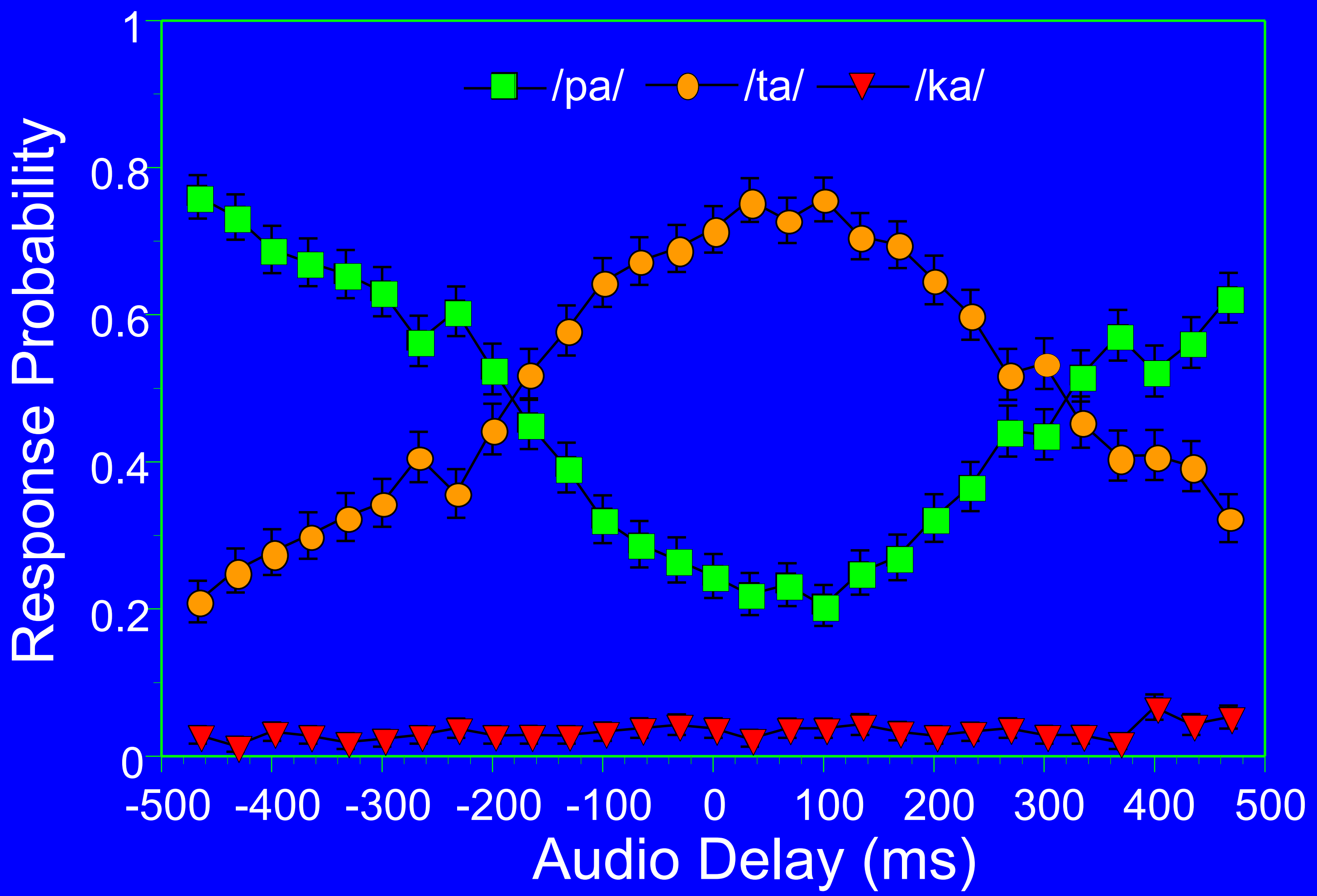
1. McGurk Illusion - Effect of A/V Asynchrony
   - $A_B V_G$ --> D or ð
   - $A_M V_D$ --> N
   - $A_P V_K$ --> T
   - $A_V V_D$ --> Z
2. Speech in Noise

McGurk Synchrony Paradigm



Demonstration of McGurk Illusion

Temporal Integration in the McGurk Effect

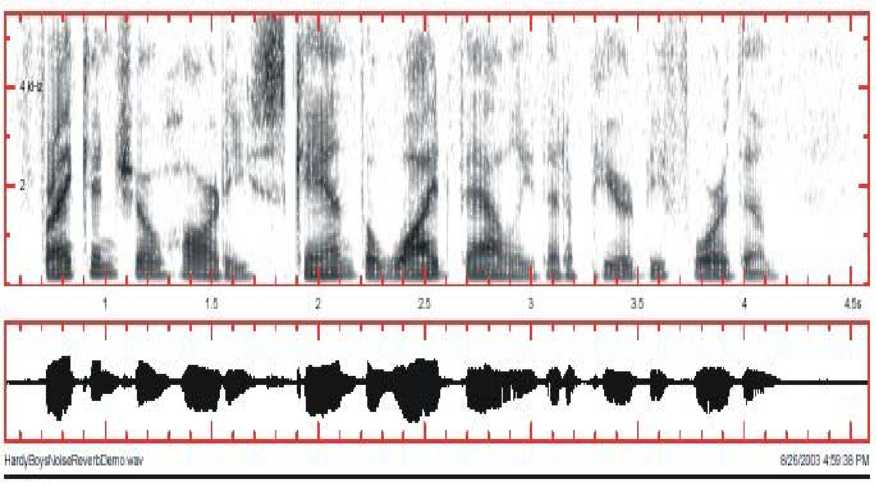


Speech Intelligibility

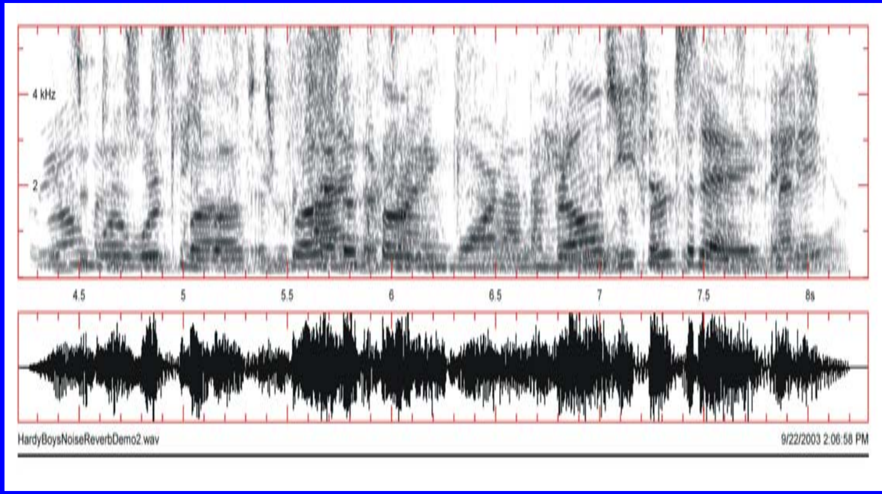
- **Number one complaint is noise and reverberation (cocktail party effect)**
- **Visual speech cues (speechreading) effectively reduce the noise by approximately 6-8 dB for most all normal and hearing-impaired individuals**

Noisy, Reverberant Speech: Demo

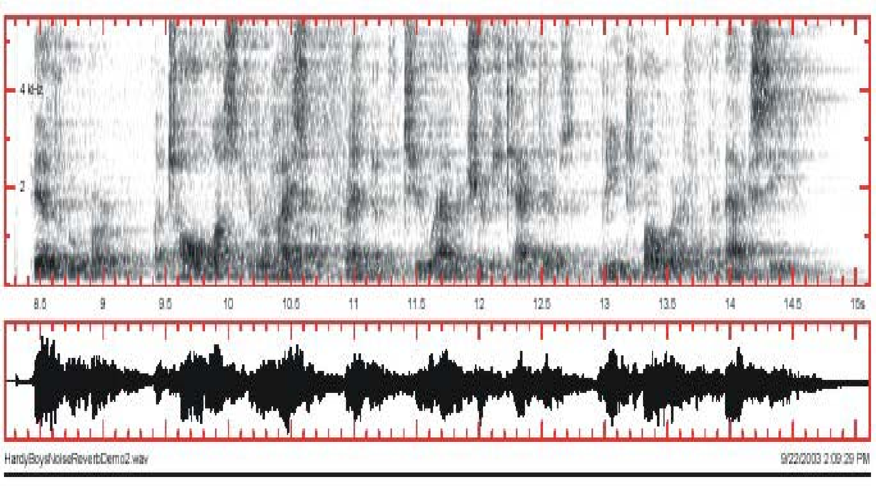
Clean Speech



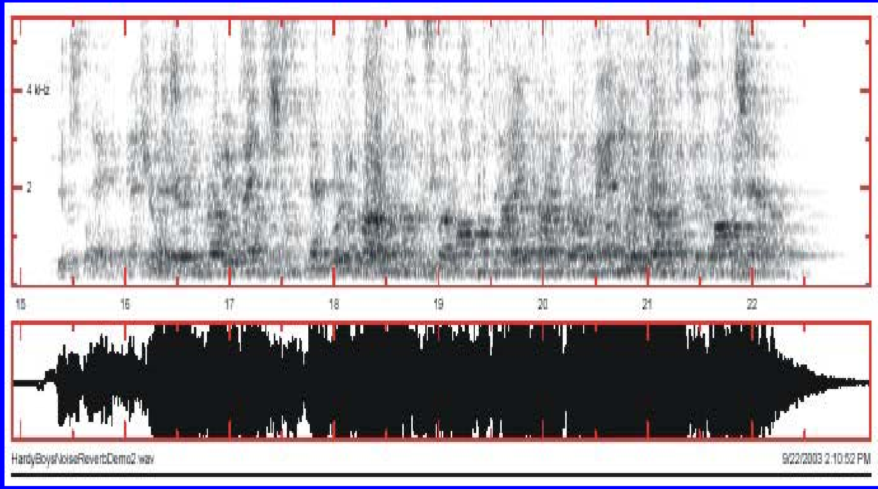
Multi-Talker Babble (4 Talkers)



Reverberation (large Conference Room)



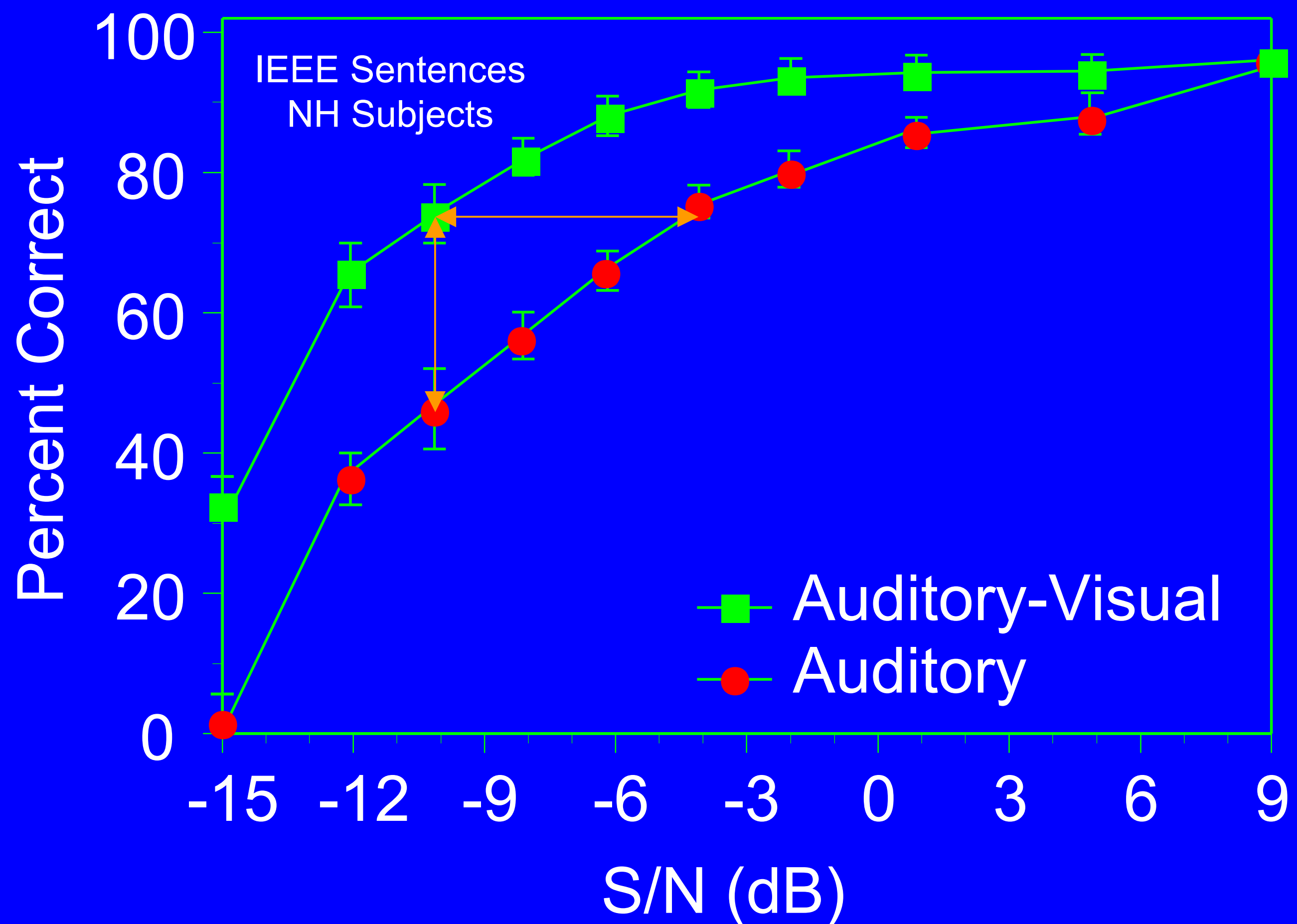
Reverberation Plus Multi-Talker Babble



Demonstration – Speech in Noise

S/N = -8 dB

Auditory-Visual vs. Audio Speech Recognition



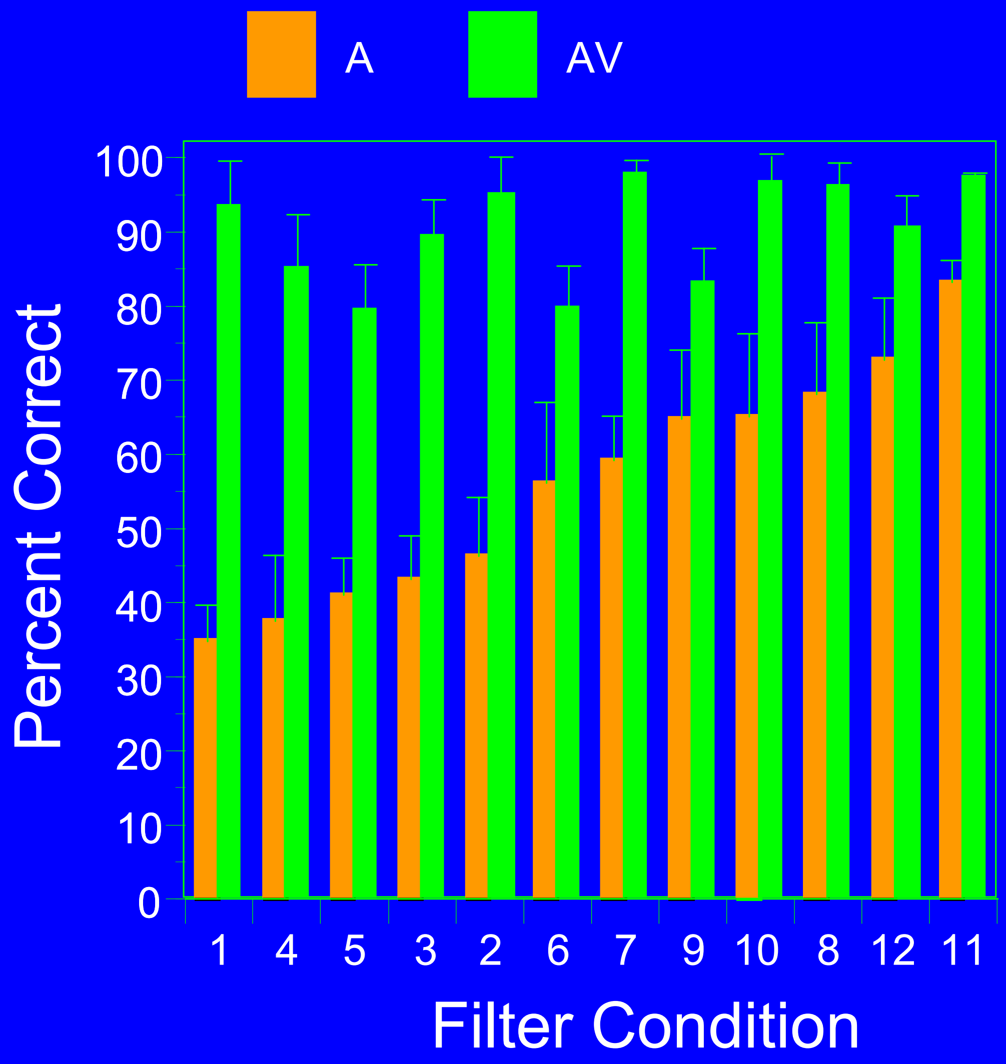
Roughly 6-8 dB improvement in S/N; roughly 30% improvement in intelligibility

Spectral Interactions

Audio-visual benefit depends on the spectral locus of the acoustic signal – <u>Visual cues are not simply additive</u>

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

Auditory-Visual Spectral Interactions: Consonants



Condition (Hz)

1. 250-505
2. 645-955
3. 1130-1515
4. 1720-2140
5. 2600-3255
6. 4200-5720
7. 250-795
8. 1130-1930
9. 3255-5720
10. 250-1130
11. 1130-2355
12. 2600-5720

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

Complementary Auditory-Visual Cues

Linguistic feature contributions to visual speech recognition. The top row represents typical feature classifications for speechreading alone (visemes). Each subsequent row represents the effects of adding information about another linguistic feature via an additional input channel (in this case auditory). Note that as additional features are added, consonant confusions associated with speechreading are resolved to a greater and greater extent.

Speechreading	p,b,m t,d,n g,k f,v θ,ð,s,z ʃ,tʃ,dʒ,ʒ l r w j
+	
Voicing	p b,m t d,n g k f v θ ð s z ʃ,tʃ dʒ,ʒ l r w j
+	
Nasality	p b m t d n g k f v θ ð s z ʃ,tʃ dʒ,ʒ l r w j
+	
Affrication	p b m t d n g k f v θ ð s z ʃ tʃ dʒ ʒ l r w j

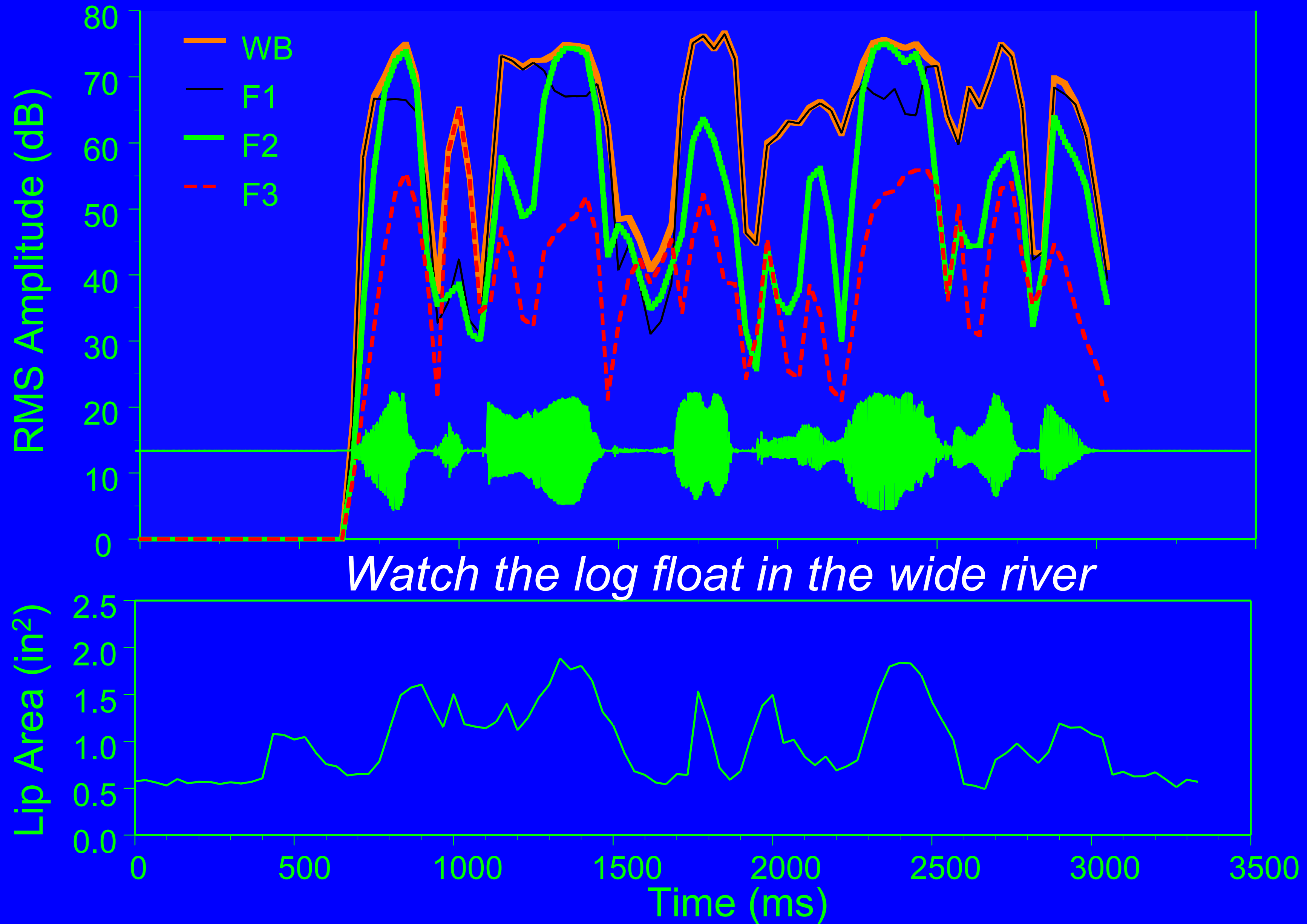
Speech Event Detection

AV Spectro-Temporal Coherence

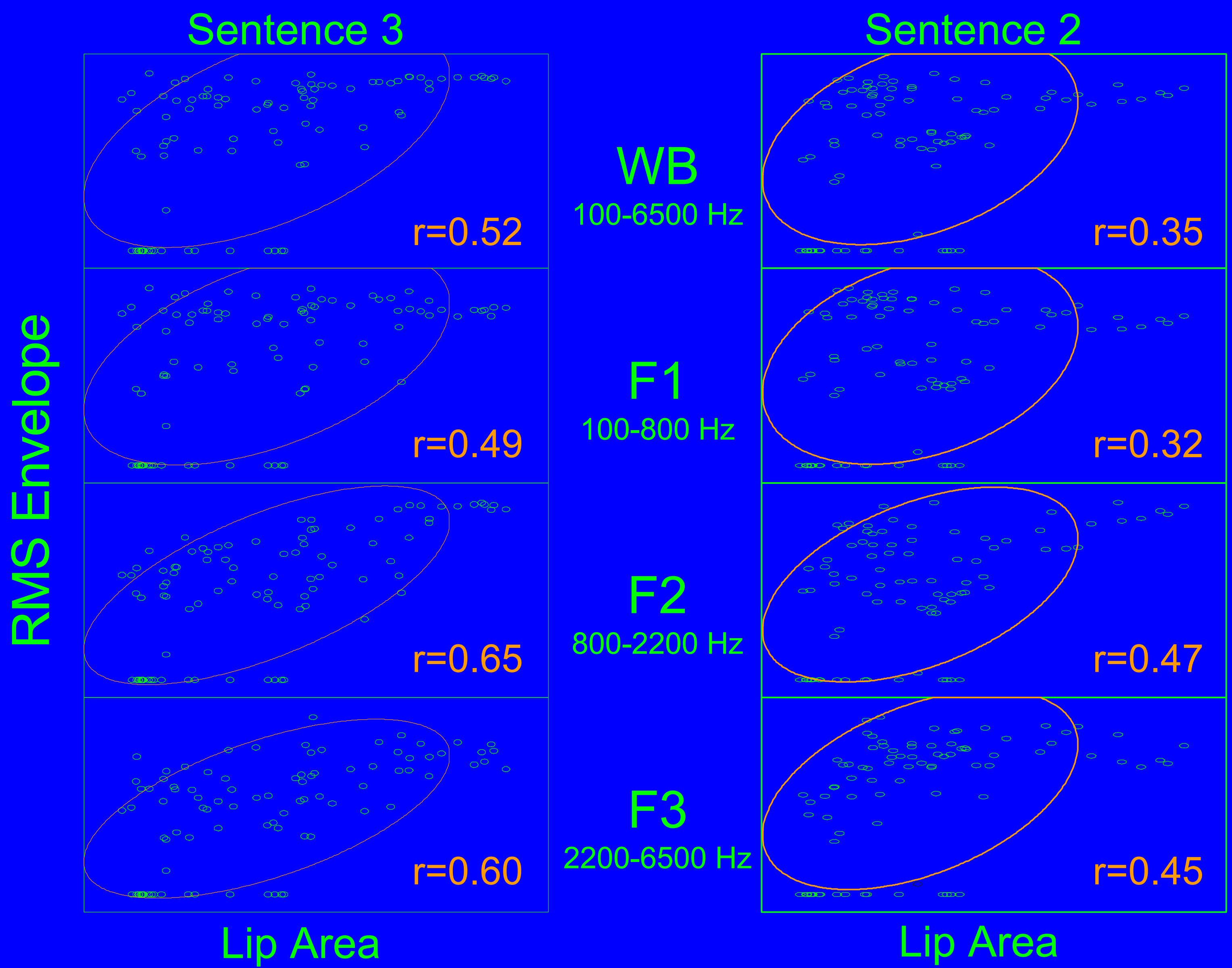
Visible articulatory kinematics are correlated with acoustic envelope (Grant and Seitz, 2000)

Degree of correlation depends on the spectral band (highest correlation found for mid-frequency bands in the F2-F3 region (Grant, 2001)

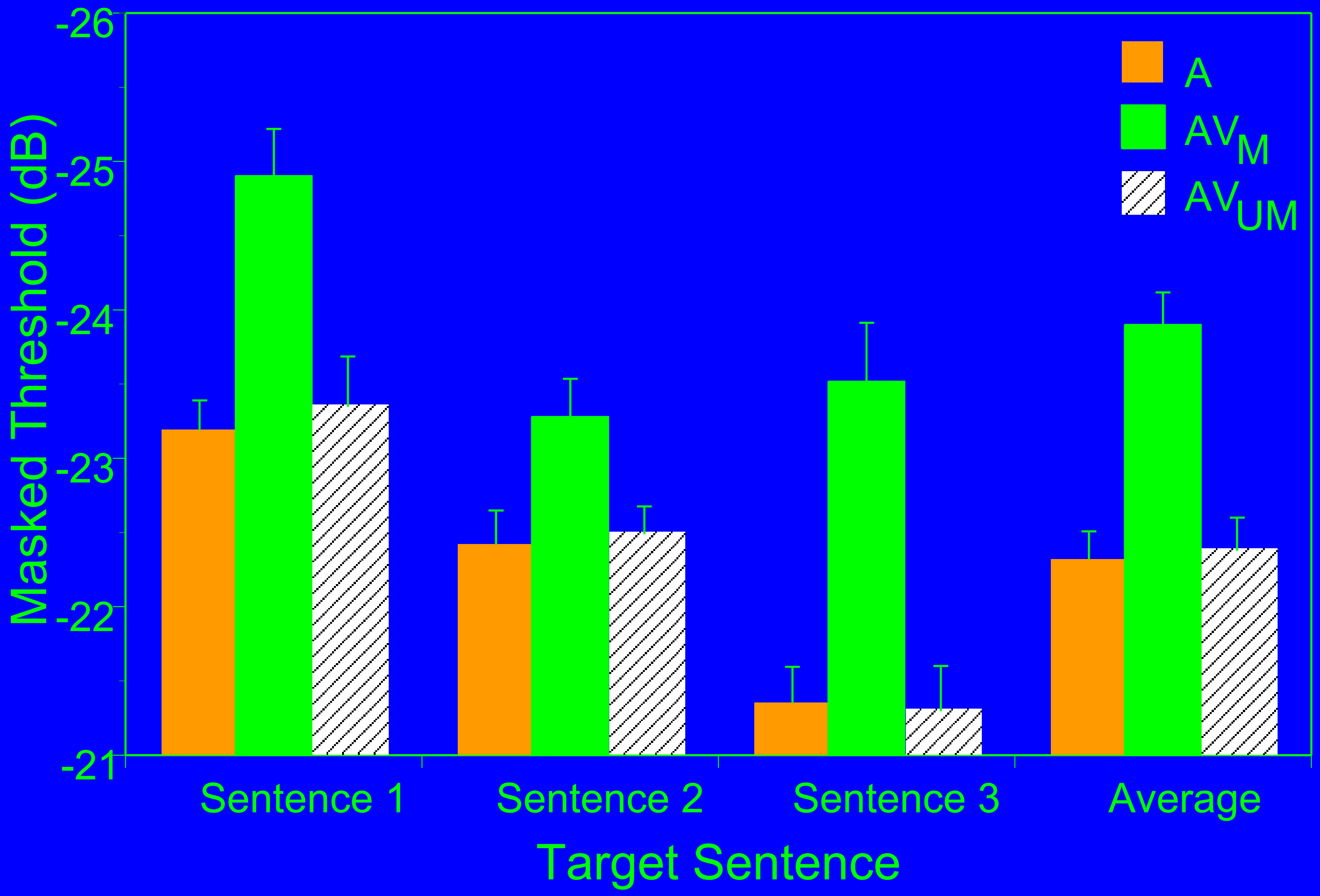
Acoustic Envelope and Lip Area Functions



Cross Modality Correlation - Lip Area versus Amplitude Envelope

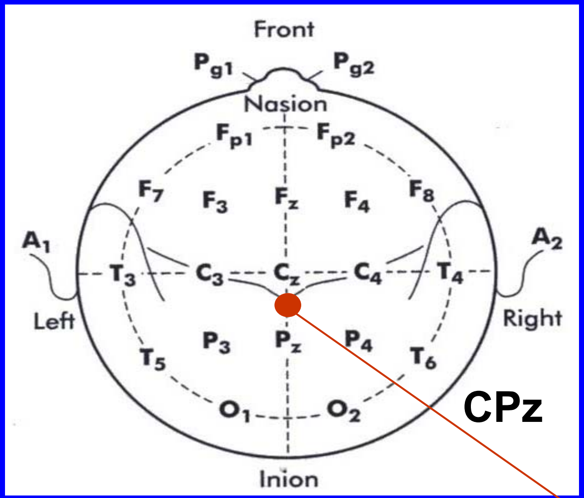


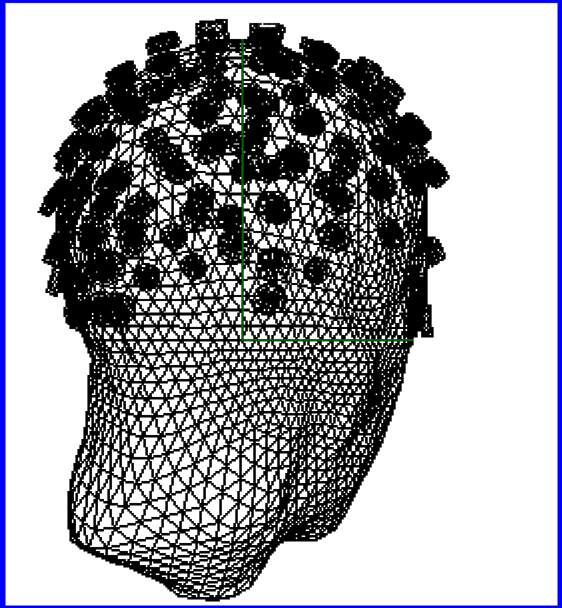
Congruent versus Incongruent Speech

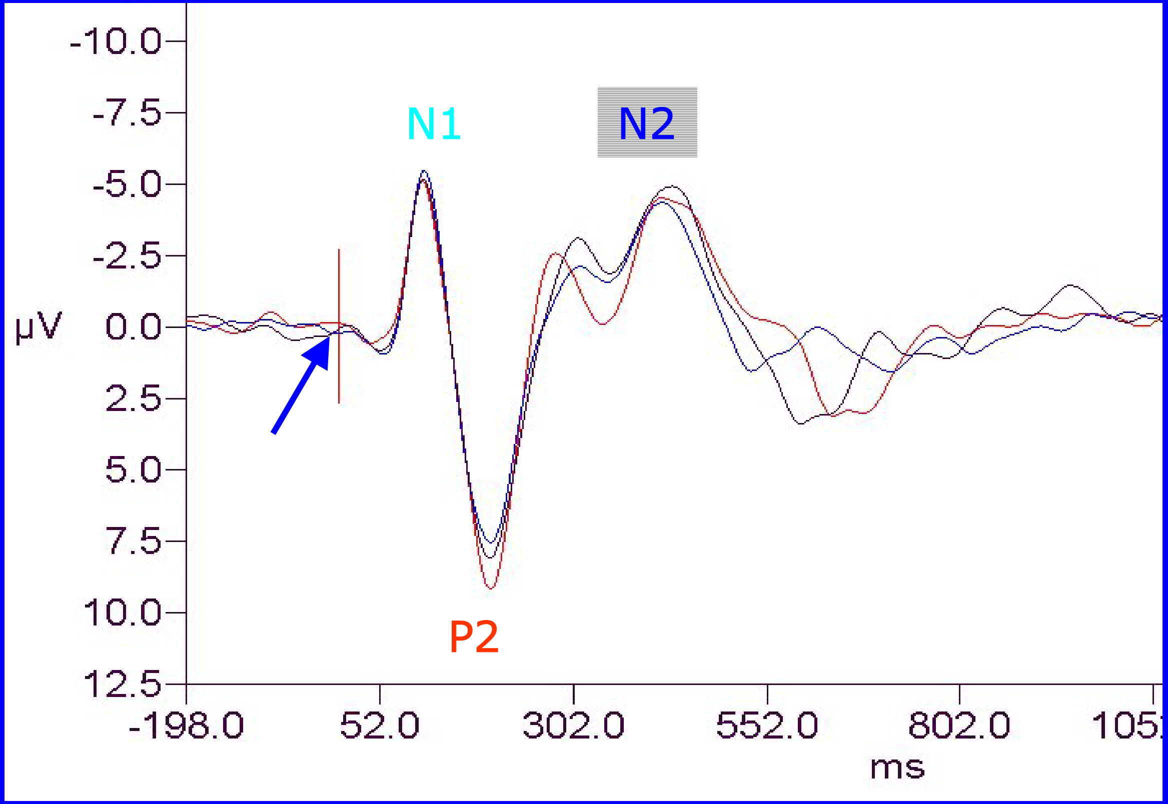


Electrophysiology

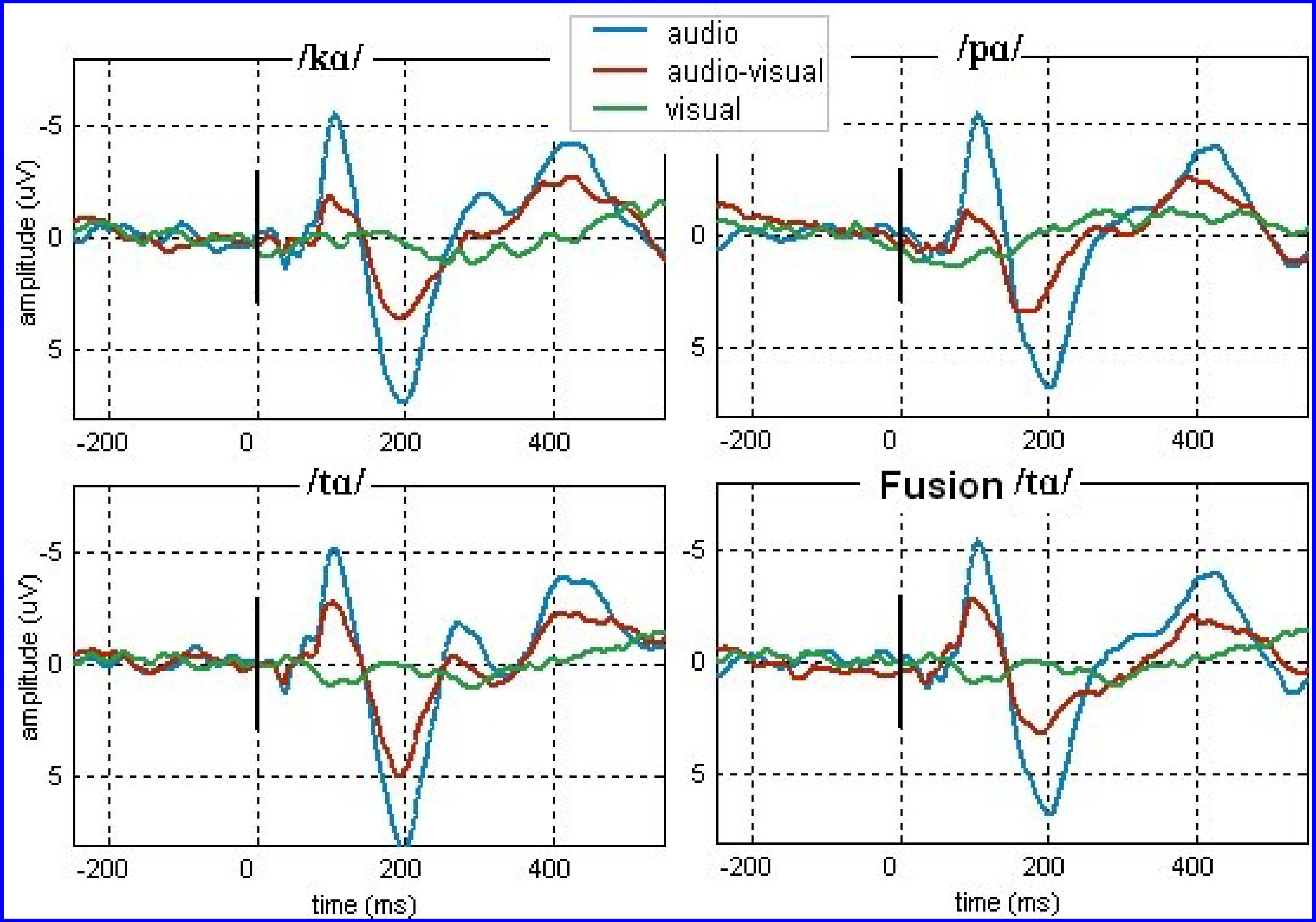
Auditory Evoked Potential (AEP) 'N1/P2 auditory complex'





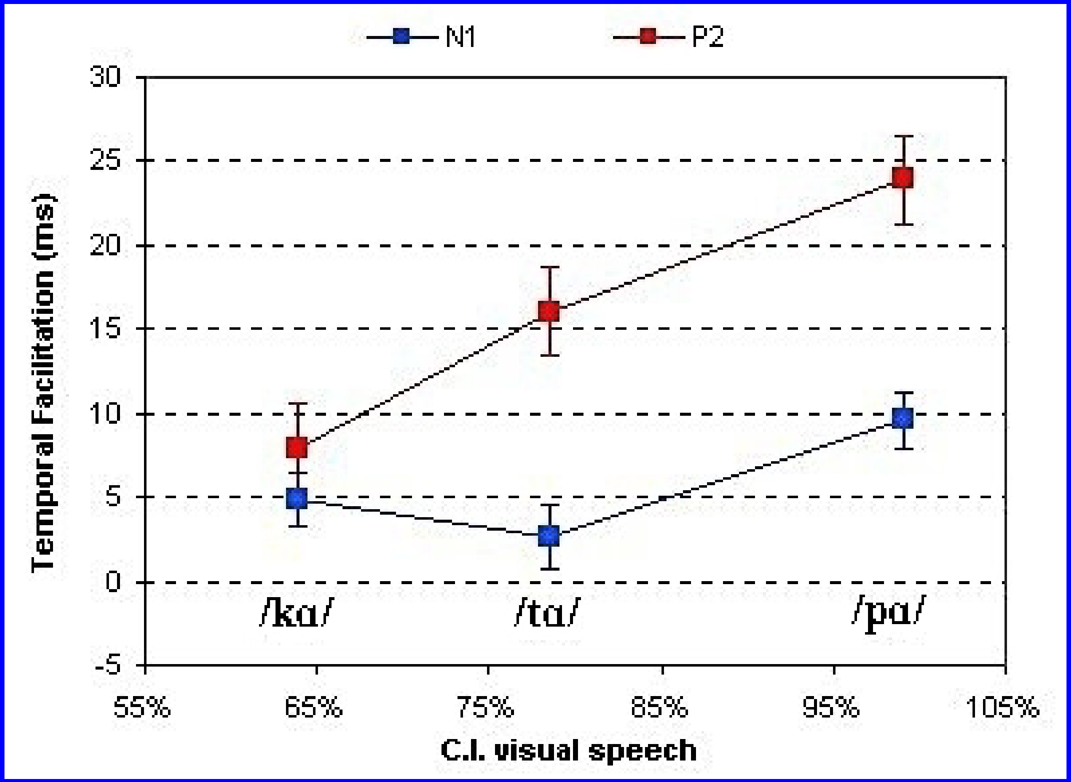


CPz - Averaged AEPs (n=16)



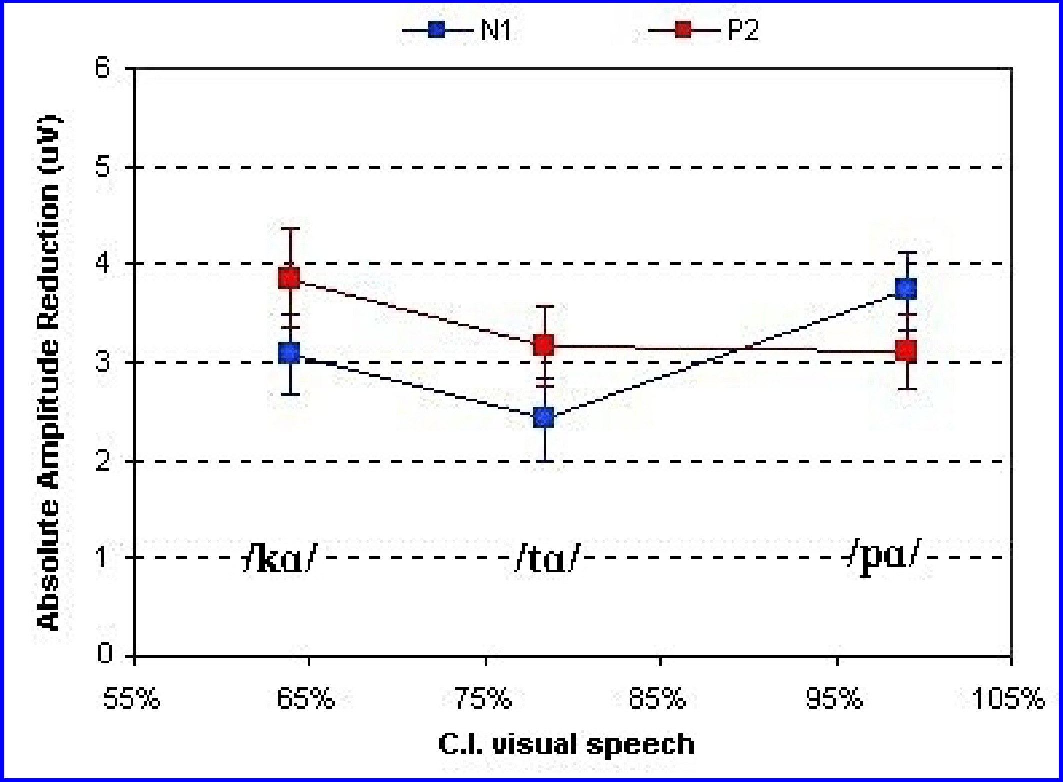
Latency Difference (A – AV) of AEPs (N1, P2) as a Function of Visual Speech Correct Identification

(A-AV) > 0 i.e.
AV occurs earlier than A



The more salient visual speech is, the faster the auditory speech processing

Amplitude Reduction (A-AV) of Early AEPs as a Function of Visual Speech Correct Identification



The AEP amplitude reduction is independent of visual speech saliency

Acknowledgments

Collaborations:

David Poeppel

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

Steven Greenberg

The Speech Institute, Oakland, CA

Funding:

NIH Grant: DC 00792-01AI

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

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